

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 02:58:27 ; Search time 256 Seconds
(without alignments)
9299.265 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atggtactcttgccacaaaaa.....ctcaagacgataatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/prodata/1/pubpna/US05_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1644	100.0	1644	10	US-09-998-284-1
2	1279.2	77.8	1801	10	US-09-824-053-30
3	42.6	2.6	1647	10	US-09-770-149-554
4	41.8	2.5	1584	9	US-09-938-842A-1942
5	41.8	2.5	1397	9	US-10-170-656-3
6	41.4	2.5	1620	9	US-09-938-842A-2266
7	40.2	2.4	1599	9	US-09-938-842A-281
8	39.2	2.4	1757	9	US-09-258-031B-72
9	38.6	2.3	1537	9	US-09-258-031B-70
10	38.6	2.3	1584	9	US-09-938-842A-1765
11	37.2	2.3	1599	9	US-09-938-842A-1762
12	36.4	2.2	1861	10	US-09-801-368-187
13	36.4	2.2	1589	9	US-09-258-031B-19
14	36.4	2.2	1784	9	US-09-258-031B-15
15	35.8	2.2	941	9	US-10-123-155-464
16	35.8	2.2	7766	9	US-10-222-162-3
17	35.8	2.2	7766	9	US-10-143-024-3
18	35.4	2.2	513509	9	US-09-754-853A-4
19	35	2.1	1756	9	US-10-114-043-7

C	20	35	2.1	2306	9	US-10-114-043-1	Sequence 1, Appl1
	21	34.8	2.1	292	10	US-09-924-035A-19	Sequence 19, Appl1
	22	34.4	2.1	604	10	US-09-770-149-927	Sequence 927, App
	23	34.2	2.1	261	9	US-09-258-031B-45	Sequence 45, Appl1
	24	34.2	2.1	1830121	9	US-10-329-960-1	Sequence 1, Appl1
	25	34	2.1	377	9	US-09-258-031B-42	Sequence 42, Appl1
	26	34	2.1	377	9	US-09-258-031B-43	Sequence 43, Appl1
	27	34	2.1	408	10	US-09-878-574-18	Sequence 18, Appl1
	28	34	2.1	695	9	US-09-258-031B-24	Sequence 24, Appl1
	29	34	2.1	7038	9	US-10-239-676-203	Sequence 203, App
	30	33.8	2.1	1981	9	US-09-258-031B-57	Sequence 57, Appl1
	31	33.6	2.0	1593	9	US-09-938-842A-319	Sequence 319, App
	32	33.4	2.0	596	10	US-09-864-761-8758	Sequence 8758, Ap
	33	33.4	2.0	1887	9	US-09-738-626-2031	Sequence 2031, Ap
	34	33.4	2.0	3309400	9	US-09-738-626-1	Sequence 1, Appl1
	35	33.2	2.0	1458	9	US-10-047-412A-3	Sequence 3, Appl1
	36	33.2	2.0	1536	9	US-09-938-842A-222	Sequence 222, Appl1
	37	33	2.0	490	12	US-10-071-751-41	Sequence 41, Appl1
	38	33	2.0	3172	9	US-09-954-531-158	Sequence 158, App
	39	33	2.0	3172	9	US-09-954-531-1580	Sequence 580, App
	40	32.8	2.0	416	9	US-09-764-868-173	Sequence 173, App
	41	32.8	2.0	716	9	US-10-123-155-512	Sequence 512, App
	42	32.8	2.0	993	9	US-09-738-626-1095	Sequence 1095, App
	43	32.8	2.0	3309400	9	US-09-738-626-1	Sequence 1, Appl1
	44	32.6	2.0	1530	9	US-09-258-031B-74	Sequence 74, Appl1
	45	32.6	2.0	1626	9	US-09-938-842A-2575	Sequence 2575, Ap

ALIGNMENTS

RESULT 1
US-09-998-284-1
; Sequence 1, Application US/09998284
; Patent No. US20020106361A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, et al.
; TITLE OF INVENTION: COMPOSITION
; FILE REFERENCE: 674509-2035
; CURRENT APPLICATION NUMBER: US/09/998, 284
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: PCT/IB00/00829
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: GB 9913050.2
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Chondrus crispus
US-09-998-284-1

Query Match 100.0%; Score 1644; DB 10; Length 1644;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCGTACTTTGGCCACAAAGACCCAGGTTACATTGTTATGACGTCACGCTGTACT	60
DB	1	ATGCGTACTTTGGCCACAAAGACCCAGGTTACATTGTTATGACGTCACGCTGTACT	60
QY	61	CCAGACAAAGCCGACCCAGATTTGCATCATGAAACAGGTTTCAACGAAGATGATT	120
DB	61	CCAGACAAAGCCGACCCAGATTTGCATCATGAAACAGGTTTCAACGAAGATGATT	120
QY	121	GGTACCAATGATTTGCTTACGTCGTTACCTCCAAAGTCCGTTACTGCTTTG	180
DB	121	GGTACCAATGATTTGCTTACGTCGTTACCTCCAAAGTCCGTTACTGCTTTG	180
QY	181	GACAGAGCTATGGAAGGTTCTCCAGGTACCGTGAAGATGTTCTGTGCTACTGT	240
DB	181	GACAGAGCTATGGAAGGTTCTCCAGGTACCGTGAAGATGTTCTGTGCTACTGT	240

241 TACGAGACTTCTTTTGAAGATGTGCAAGGCTATTATCAAGTTACTGTTGGTT 300
241 TAGGAAACCTTCGTTTGAAGATGTGCAAGGCTATTATCAAGTTACTGTTGGTT 300
301 GAATCTGGTTACGACGAGATAGAGTTTACTGTTCTTCCGTTGACACCACTGGGGT 360
301 GAATCTGGTTACGACGAGATAGAGTTTACTGTTCTTCCGTTGACACCACTGGGGT 360
361 TCCCTTCAAGACCTTGTTCAGAGACCAAGGTTTGTGACAGGTTGCTGTTACTCC 420
361 TCCCTTCAAGACCTTGTTCAGAGACCAAGGTTTGTGACAGGTTGCTGTTACTCC 420
421 GTGCGTTGGGTGTGTCATGTCGTGAGGTGACGATTTTGGCCAGATTGACAGGT 480
421 GTGCGTTGGGTGTGTCATGTCGTGAGGTGACGATTTTGGCCAGATTGACAGGT 480
481 TTGCGACTGATTTGTTTATCCGTTGTTGAACTTGTGTTAAGCCAGTCTTGAACG 540
481 TTGCGACTGATTTGTTTATCCGTTGTTGAACTTGTGTTAAGCCAGTCTTGAACG 540
541 TCTGTTCTTAAGTACGTTTCAAGGATTTCCGAGGTAAACGAGTGTGTTGGGCT 600
541 TCTGTTCTTAAGTACGTTTCAAGGATTTCCGAGGTAAACGAGTGTGTTGGGCT 600
601 CACACTGTGAGGTGAGGTAACTTGTGTTTATCAACCAATCTACTTCAAGATTTG 660
601 CACACTGTGAGGTGAGGTAACTTGTGTTTATCAACCAATCTACTTCAAGATTTG 660
661 CCAATGTTCTCAAGAGGTGTGATGCTTCTTCACTTCACTTCTTGGACGTTTCACT 720
661 CCAATGTTCTCAAGAGGTGTGATGCTTCTTCACTTCACTTCTTGGACGTTTCACT 720
721 AGAGATGCTTGAAGATTTGTTGACTAAGTACTTCAAGTGTGCTAGATTTGAG 780
721 AGAGATGCTTGAAGATTTGTTGACTAAGTACTTCAAGTGTGCTAGATTTGAG 780
781 AATACGTGTGTAAGTTCCAATCTTCCACCAACAGTGAAGATTTGTTATGTA 840
781 AATACGTGTGTAAGTTCCAATCTTCCACCAACAGTGAAGATTTGTTATGTA 840
841 TATACATCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
841 TATACATCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
901 GAGGCTGACATTTGAACGAGATTTCAAAACATGAGGCTTCAAGGCTTGGGTCAT 960
901 GAGGCTGACATTTGAACGAGATTTCAAAACATGAGGCTTCAAGGCTTGGGTCAT 960
961 GCTGGTGGGCTCTTTCCTGTTGACCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1020
961 GCTGGTGGGCTCTTTCCTGTTGACCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1020
1021 ATGCTATGACGAGACTATGAGACTTCTTCTTCAAGGCTTGAAGAGAGAGAGAG 1080
1021 ATGCTATGACGAGACTATGAGACTTCTTCTTCAAGGCTTGAAGAGAGAGAGAG 1080
1081 GGTCTTATCAAG 1140
1081 GGTCTTATCAAG 1140
1141 CAGATTGATTTATCTGAAATACCTTACTAGAGTTCTGAGGTTTGAAGTGGGAA 1200
1141 CAGATTGATTTATCTGAAATACCTTACTAGAGTTCTGAGGTTTGAAGTGGGAA 1200
1201 ATGAGAGATGCTTCTTCAAGGTTGATGTTCCGTTGAGATTTCAAGGTTGTTGG 1260
1201 ATGAGAGATGCTTCTTCAAGGTTGATGTTCCGTTGAGATTTCAAGGTTGTTGG 1260
1261 GATGCTCTGAGTGTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1261 GATGCTCTGAGTGTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1321 CAG 1380

1321 CAG 1380
1381 TATGAGCTTATGAGTGTGTTCCAGACCTTAAACATCAAGTGTGAGTGAAGGTT 1440
1381 TATGAGCTTATGAGTGTGTTCCAGACCTTAAACATCAAGTGTGAGTGAAGGTT 1440
1441 TTTGAGGAGATGCTTCACTTCACTACCTGATGTTGATGATGATGATGATGATG 1500
1441 TTTGAGGAGATGCTTCACTTCACTACCTGATGTTGATGATGATGATGATGATG 1500
1501 TATGAGCTTATGAGTGTGTTCCAGACCTTAAACATCAAGTGTGAGTGAAGGTT 1560
1501 TATGAGCTTATGAGTGTGTTCCAGACCTTAAACATCAAGTGTGAGTGAAGGTT 1560
1561 TTTGAGGAGATGCTTCACTTCACTACCTGATGTTGATGATGATGATGATGATG 1620
1561 TTTGAGGAGATGCTTCACTTCACTACCTGATGTTGATGATGATGATGATGATG 1620
1621 GAGCTTAAAG 1644
1621 GAGCTTAAAG 1644

RESULT 2
US-09-824-053-30
Sequence 30, Application us/09824053
Patent No. US20020106725A1
GENERAL INFORMATION:
APPLICANT: Peter Stougaard
Ole Cal Hansen
TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hulton & Williams
STREET: 1900 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006-1109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/824,053
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/669,304
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Stanislaus Akseman
REGISTRATION NUMBER: 28,562
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 955-1926
TELEFAX: (202) 778-2201
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 84..1721
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-824-053-30

Query Match 77.8%; Score 1279.2; DB 10; Length 1801;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
QY 1 ATGGCTACTTTGGCCAAAGAGCCAGGTTACATTTGATGACGTCAACGCTGACT 60
DB 84 ATGGCTACTTTCTCAGAAAGACCCGGTTATATGTAATGATGTCAACGGGGCACCC 143
QY 61 CCAGACAAACCCGACCCAGCTCTCCCTCATGAAACAGGGCTTCAACGGCGCTGGATT 120
DB 144 GGGGAAACCCGACCCAGCTCTCCCTCATGAAACAGGGCTTCAACGGCGCTGGATT 203
QY 121 GGTAAACAATCGATTTGCTTTACGTGTTTACATCCCAAGGCTCTGTAAGCTTTG 180
DB 204 GGAACCTAATATGATTTGCTTTATGTCGTGTAACCTCCAGAGTCTTGTAAGCTT 263
QY 181 GACAGAGCTATGAAAGTGTCTCCAGGTACCGTCAGAAATGCTTTCTGGTGTCACTGT 240
DB 264 GACCGGTATATGAAAGTGTCTCCCGGTACAGTCAGATCGTCTCTGGCGGCAATGCG 323
QY 241 TACGAAGACTTGGTTTCCAGCAATGTGTCAAGGCTATATCAAGTTACTGTTGGTT 300
DB 324 TACGAGAGCTTGGTATTTGACGATGCGTCAGAGCCATCATCAAGCTGCTGCTGCT 383
QY 301 GAATCTGTTTACGACGACGATAGAGGTTACTTCTCTCCGCTGACCAACTGGGGT 360
DB 384 GAGAGGTGTTATATACACATAGGGGTTACTTCTGACAGCTGGAATCAAAATTTGGGG 443
QY 361 TCCCTTCAAGACCTTGTTCAGAGACCAACGATGATTTTGGCAGTGGTCTGTCTCC 420
DB 444 TCCCTTCAAGACCTTGTTCAGAGACCAACGATGATTTTGGCAGTGGTCTGTCTCC 503
QY 421 GTGCGTTTGGGTGTCACTTGTGCGTGAAGGTGAAGGATTTTGGCAGATTTGCAAGGT 480
DB 504 GTGCGGCTGCGTGGCCACATTTGCGGAGGTGAAGGATTTTGGCAGATTTGCAAGGT 563
QY 481 TTGCGAGCTGATTTGTTATCCGCTGTTGAAGTGTGCTTAAAGCAGCTTGAACGAGAC 540
DB 564 CTCCCGCTGATTTGGCTCAGCGGCGTGGAGGTGCTGTTAAGCAGCTTGAACGAGAC 623
QY 541 TCTGTTCTTAAGTACGTTCAAGAGATTTCCGAAGGTGAACGAGGTGATTTTGGGCT 600
DB 624 TCGGTACTAAGTATGTGCAAAAGATTTCCGAAGGTGAACGAGGTGATTTTGGGCA 683
QY 601 CACACTGTGTGAGGTGAGGTAACTTGGTATTTATCAACCAATATCTTCAAGATTTG 660
DB 684 CACACAGGTGTGCGGTGCGGAACTTTGGAATCATCAACCAATATCTTCAAGATTTG 743
QY 661 CCAATGTTCCAGAGGTGTCACTGCTTCACTTCACTTCTCTGGGAGGGTTTCACT 720
DB 744 CCAATGTTCTCAGAGGTGTCACTGCTTCACTTCACTTCTCTGGGAGGGTTTCACT 803
QY 721 AGAGATGCTTGGCAAGATTTTGTGACTAATGACTTCAAGTTGCTAGATGATTTGAG 780
DB 804 AGAGATGCTTGGCAAGATTTTGTGACTAATGACTTCAAGTTGCTAGATGATTTGAG 863
QY 781 AATACGTGTGTGATTTTCAATTTTCAACCAAGCAGCTGAAGATTTGTTATGTA 840
DB 864 AATACGTGTGTGATTTTCAATTTTCAACCAAGCAGCTGAAGATTTGTTATGTA 923
QY 841 TATACATCTCTACTCTAAGAGCGCGGAGAGAAAGTTGCGCAAGACAGACATATCATTTG 900
DB 924 TATACATCTCTACTCTAAGAGCGCGGAGAGAAAGTTGCGCAAGACAGACATATCATTTG 983
QY 901 GAGGCTGACATTTGAACAGATCTACAAAGACGAGGCTTCAAGGCTTGTGTGTGAT 960
DB 984 GAGGCTGACATTTGAACAGATCTACAAAGACGAGGCTTCAAGGCTTGTGTGTGAT 1043
QY 961 GCTGTGTGTGTGCTTCTTCTGTTTGAACCTAAGAAAGACACATCTCAAGATTTTAT 1020
DB 1044 GCTGTGTGTGTGCTTCTTCTGTTTGAACCTAAGAAAGACACATCTCAAGATTTTAT 1103
```

```
QY 1021 ATGCATGACGACATATGACTACCCCTTTCTAAGCTTTGACTGAGACATATCAACGCTTCC 1080
DB 1104 ATGCATGACGACATATGACTACCCCTTTCTAAGCTTTGACTGAGACATATCAACGCTTCC 1163
QY 1081 GGTCTTAATCAGAGAGGTATGATCAAGTCTGTTTACATGATCAAGATCTTTCAGACTTC 1140
DB 1164 GGTCTTAATCAGAGAGGTATGATCAAGTCTGTTTACATGATCAAGATCTTTCAGACTTC 1223
QY 1141 CAGATTTGATTTATCTGAAATATACCTTACTAGGTTTCTGACGTTTGTGACTAGTCCGAA 1200
DB 1224 CAGATTTGATTTATCTGAAATATACCTTACTAGGTTTCTGACGTTTGTGACTAGTCCGAA 1283
QY 1201 ATGAAGATGCTCTTCTTCCAGGTTATATGTTTGGTGTGATTTCAAGGTTGTTGG 1260
DB 1284 ATGAAGATGCTCTTCTTCCAGGTTATATGTTTGGTGTGATTTCAAGGTTGTTGG 1343
QY 1261 GATGCTATGCTGAGTGTCTTACAGAGATATCATCAAACTGCAATGACCAATATCTGG 1320
DB 1344 GATGCTATGCTGAGTGTCTTACAGAGATATCATCAAACTGCAATGACCAATATCTGG 1403
QY 1321 CAGAAAGACAGAGATGACGTTAACTTGAAGTGAATTAGAGACTTTTACAGAGAGATG 1380
DB 1404 CAGAAAGACAGAGATGACGTTAACTTGAAGTGAATTAGAGACTTTTACAGAGAGATG 1463
QY 1381 TATGAGCTTATGTTGTGTTTCAAGCTTAACTCAAGTGTGAGTGAAGTGAAGTGT 1440
DB 1464 TATGAGCTTATGTTGTGTTTCAAGCTTAACTCAAGTGTGAGTGAAGTGAAGTGT 1523
QY 1441 TTTGAGGATGCTATCTTCACTTACCTGATGTTGATTTGAACAATGGAAGACGTTAAG 1500
DB 1524 TTTGAGGATGCTATCTTCACTTACCTGATGTTGATTTGAACAATGGAAGACGTTAAG 1583
QY 1501 TATGAGCTTATGTTGTGTTTCAAGCTTAACTCAAGTGTGAGTGAAGTGAAGTGT 1560
DB 1584 TATGAGCTTATGTTGTGTTTCAAGCTTAACTCAAGTGTGAGTGAAGTGAAGTGT 1643
QY 1561 TTTGAGGATGCTATCTTCACTTACCTGATGTTGATTTGAACAATGGAAGACGTTAAG 1620
DB 1644 TTTGAGGATGCTATCTTCACTTACCTGATGTTGATTTGAACAATGGAAGACGTTAAG 1703
QY 1621 GAGCTTAAAGACATTAATATGATG 1644
DB 1704 GAGCTTAAAGACATTAATATGATG 1727

RESULT 3
US-09-770-149-554/c
; Sequence 554, Application US/09770149
; Patent No. US200205963A1
; GENERAL INFORMATION:
; APPLICANT: Gortlach, John
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesener, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurlan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
```

;; CURRENT FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: 60/178,506
;; PRIOR FILING DATE: 2000-01-27
;; NUMBER OF SEQ ID NOS: 999
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 554
;; LENGTH: 647
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(647)
;; OTHER INFORMATION: n = A,T,C or G
US-09-770-149-554

Query Match 2.6%; Score 42.6; DB 10; Length 647;
Best Local Similarity 62.9%; Pred. No. 0.0047;
Matches 66; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1519 TACTTTTGGGTAACTGACAGATTGATCAAGCGCAATGTTGGATCTTAACGAG 1578
DB 253 TATTCTTGGGAATTGGAAGATTGATGATCTTAAGCTAGATGATCTGATAT 134

QY 1579 ATCTTCAACAACAACGCTATCTCTACTAACTCTTAAAGAG 1623
DB 193 TTCTTCAAGAAGCAGACAGATATCTCCGGTTCGTGTAAGTAG 149

RESULT 4
US-09-938-842A-1942
; Sequence 1942, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1942
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1942

Query Match 2.5%; Score 41.8; DB 9; Length 1584;
Best Local Similarity 50.2%; Pred. No. 0.016;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 315 CGACGATGAGGTTACTTCTCTCCGGTGACACCAACTGGGTTCTTCAAGACCTT 374
DB 417 CGAGAGAACGGCTGGTGTGATCTGGTGCTACCTCGAGAGCTTACTATGAATCTC 476
QY 375 GTTCAGAGACCAAGTGTGAGTTTGCAGGTGTTCTGTACTCCGCGTTGGGTGG 434
DB 477 TGAGAGAGCAATGTTCTTGATTTCCGGCGGTTGTCTTACCAACATGGCGGTGGTGG 536
QY 435 TCACATTGTGCGTGGAGTGAAGCGTATTTTGGCAGATTGACAGGTTTGCAGTGAATTG 494
DB 537 ACACTTAGCGGCGGAGAGATACGTAATCTGATGAGAAAGATGTTTGTCTGCGTGAATA 596
QY 495 GTTATCCGGTGTGAAGTGTGCTT 519

DB 597 CGTTTCCGCTCCGGGATCGTTGAT 621

RESULT 5
US-10-170-656-3
; Sequence 3, Application US/10170656
; Publication No. US20030028927A1
; GENERAL INFORMATION:
; APPLICANT: Lindquist, Susan K.
; APPLICANT: Quetesch, Christine
; APPLICANT: Sangster, Todd A.
; TITLE OF INVENTION: Methods and Compositions for Revealing Hidden Generic Variation in
; FILE REFERENCE: ARCD:377U51
; CURRENT APPLICATION NUMBER: US/10/170,656
; CURRENT FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/298,211
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/379,484
; PRIOR FILING DATE: 2002-05-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2397
; TYPE: DNA
; ORGANISM: wheat
US-10-170-656-3

Query Match 2.5%; Score 41.8; DB 9; Length 2397;
Best Local Similarity 52.6%; Pred. No. 0.022;
Matches 91; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1409 CTACACTGAGTGTGAGTGTGAAGTGTGTTTGGAGGATCTTCACTACCTTG 1468
DB 768 CTGACGATGAAGTGAAGTGAAGTGAAGATCTGATGAGGCGCAAGTTGACGAATTG 827
QY 1469 ATCTTGAATTGAACAACCTGGAGAACGTTAAGTATGTGCTTGGAACTTACTTTTGG 1528
DB 828 ATGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 887
QY 1529 GTAACTGGAACGATGATCAAGGCCCAATGTTGGGATCTTCAAGATC 1581
DB 888 GGAACCTGATCAACAAGCAGAGGCCCATCTTGATGAGAGAGCTTGAGAGATC 940

RESULT 6
US-09-938-842A-2266
; Sequence 2266, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2266
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2266

Query Match 2.5%; Score 41.4; DB 9; Length 1620;
Best Local Similarity 53.4%; Pred. No. 0.022;

Matches	87	Conservative	0	Mismatches	76	Indels	0	Gaps	0
QY	357	GGTTTCCTTCAAGACCTTGTTCAGAGACCA	CGGTAGAGTTTGGCAGGTGTTCTCTGTTA	416					
Db	486	GGTTATATTACAGAAATTTGGAGAAAAGCAAAAGTCATGATTTCCCGCGGAGTTTGTC	545						
QY	417	CTCCGTCCGTTTGGGTGTCACATTGTCCGTGAGAGTACCGTATTTTGGCCAAATTGCA	476						
Db	546	CACGGTGTGTTGGTGACATATPAAAGGGGTGGTTACCGTATATATGTTGAGAAATT	605						
QY	477	CGGTTTGCACGATGGTTATCCGGGTGTGAAGTTGCGTT	519						
Db	606	CGGATTTATCAGTTGATTAACCTTGATGATCCGAAAGATCTGTGAT	648						

```

RESULT 7
US-09-938-842A-281
; Sequence 281, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIORITY FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIORITY FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIORITY FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIORITY FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 281
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-281

```

```

Query Match      2.4%  Score 40.2;  DB 9;  Length 1599;
Best Local Similarity 57.6%;  Pred No. 0.055;
Matches 72;  Conservative 0;  Mismatches 53;  Indels 0;  Gaps 0

Oy      393  AGTTTGGCAGATGGTCTCTTTACTCCGTCGTTTGAGTGATCAATTGTCGTGAGG 452
Db      516  AGCGTTTCCGCCCGGAGTTTGCCACCAGTCGGCGTCGGAGACATTTTAGCGGGGAGG 575

Oy      453  TGACGGTATTTTGGCAGATTGCACGGTTTGGCACTGCATTGTTATCCGTTGAACT 512
Db      576  ATACGGTACTTTCATGAGAAAAACACGGTTTAGCAGCTATCACTGATAGACGACGTGT 635

Oy      513  TGTGC 517
Db      636  TGTGC 640

RESULT 8
US-09-258-031B-70
: Sequence 70, Application US/09258031B
: Patent No. US20020168735A1
: GENERAL INFORMATION:
: APPLICANT: STUIVER, Maarten Hendrik
: APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor
: APPLICANT: SELA-BURLAGE, Marianne Beatrix
: APPLICANT: MELCHERS, Leo Sjoerd
: APPLICANT: VAN DEVENTER-TROOST, Johanna Pieterella
: APPLICANT: LAGEMEIJ, Wessel
: APPLICANT: PONSSTEIN, Anne Silene
: APPLICANT: LAGEMEIJ, Wessel
: APPLICANT: PONSSTEIN, Anne Silene

```

```

1  TITLE OF INVENTION:  ANTIUNGUAL PROTEINS, DNA COILING
2  TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
3  TITLE OF INVENTION:  SAME.
4  NUMBER OF SEQUENCES:  75
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  LADAS & PARRY
7  STREET:  26 WEST 61 STREET
8  CITY:  NEW YORK
9  STATE:  NY
10 COUNTRY:  USA
11 ZIP:  10023 - 7604
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  3.25" floppy disk
14 COMPUTER:  IBM PC compatible
15 OPERATING SYSTEM:  WINDOWS 95
16 SOFTWARE:  WORDPERFECT 8
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER:  US/09/258,031B
19 FILING DATE:  25-FEB-1999
20 CLASSIFICATION:  435
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER:  PCT/EP97/04923
23 FILING DATE:  04-SEP-1997
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER:  EP97200831.2
26 FILING DATE:  19-MAR-1997
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER:  EP96202466.7
29 FILING DATE:  04-SEP-1996
30 INFORMATION FOR SEQ ID NO:  70:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH:  1757 base pairs
33 TYPE:  nucleic acid
34 STRANDEDNESS:  double
35 TOPOLOGY:  linear
36 MOLECULE TYPE:  DNA (genomic)
37 HYPOTHEICAL:  NO
38 ANTI-SENSE:  NO
39 ORIGINAL SOURCE:
40 ORGANISM:  Arabidopsis thaliana
41 STRAIN:  Colombia
42 FEATURE:
43 NAME/KEY:  CDS
44 LOCATION:  join(1..570, 801..1754)
45 US-09-258-031B-70

```

```

Query Match      2.4%; Score 39.2; DB 9; Length 1757;
Best Local Similarity 53.2%; Pred. No. 0.13; Mismatches 73; Indels 0; Gaps 0;
Matches 83; Conservative 0;

Qy      393 AGTTTCCGAGAGTGGTTCCTGTTACTCCGTCGGTGGTGTCACATTTGCGGTGAGG 452
Db      438 AGCTTATCCCGCCGGAATTTGTCCACGGTTGGTGTGGTGGCATATCATGGTGGAGG 497

Qy      453 TGACGGTATTTTGGCCAGATTGCACGGTTTCCAGTCGATTTGCTTATCCGTTGAAGT 512
Db      498 TTACGGTAACTGATGAGAAAATACGGTCTCACCGTATATATACATTCATGACAAGAT 557

Qy      513 TGTGCTTAAGCCAGTCTTGACCCGAAGACTCTGTTCT 548
Db      558 GGTGACGTTAATGTATTAATGTATATCTTATTTT 593

RESULT 9
US-09-258-031B-72
; Sequence 72, Application US/09258031B
; Patent No. US20020168735A1
; GENERAL INFORMATION:
; APPLICANT: STUYVER, Maarten Hendrik
; APPLICANT: CUSTERS, Jérôme Humbertina Henricus Victor
; APPLICANT: SELA-BURLAGE, Marianne Beatrix
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: VAN DEVENTER-TROOST, Johanna Pietermella

```

```

; APPLICANT: LAGEMEG, Wessel
; APPLICANT: PONSSTEIN, Anne Silene
; APPLICANT: LAGEMEG, Wessel
; APPLICANT: PONSSTEIN, Anne Silene
; TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
; TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LADAS & PARRY
; STREET: 26 WEST 61 STREET
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10023 - 7604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: WORDPERFECT 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/258,031B
; FILING DATE: 25-FEB-1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP97/04923
; FILING DATE: 04-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP97200831.2
; FILING DATE: 19-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP96202466.7
; FILING DATE: 04-SEP-1996
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1527 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: Colombia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1524
; US-09-258-031B-72

Query Match
Best Local Similarity 56.8%; Pred. No. 0.18; Length 1527;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 393 AGTTTCCAGGTGCTCTGTTACTCCGTCGGTTGGTGTCACATTGTCGGTGAGG 452
    |||||
DB 438 AGCTTATCCCGCGGAATTTGTCACAGGTGGTGCGCATATCAGTGTGAGG 497
    |||||
QY 453 TGACGGTATTTGGCCGATTTGCACGCGTTGGCCAGTGGATTGCTTCCGTTGAAGT 512
    |||||
DB 498 TTACGGTAACATGATGAGAAATACGGTCTCACCGGTAGATAATATCCATCATGCAAGAAAT 557
    |||||
QY 513 TGTGG 517
    |||||
DB 558 GGTGG 562

RESULT 10
US-09-938-842A-1765
; Sequence 1765, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1762
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1762

Query Match
Best Local Similarity 50.0%; Pred. No. 0.54; Length 1599;
Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 297 GGTGAATCGGTAGACAGATAGAGTTACTGCTCTCTCGGTGACCAACTG 356
    |||||
DB 420 GGTGACGTTGATTGACAGTACAGTACAGTGTGCGTCACTGCTTACCATCGGAGA 479
    |||||
```

```

; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1765
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1765

Query Match
Best Local Similarity 56.8%; Pred. No. 0.18; Length 1584;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 393 AGTTTCCAGGTGCTCTGTTACTCCGTCGGTTGGTGTCACATTGTCGGTGAGG 452
    |||||
DB 495 AGCTTATCCCGCGGAATTTGTCACAGGTGGTGCGCATATCAGTGTGAGG 554
    |||||
QY 453 TGACGGTATTTGGCCGATTTGCACGCGTTGGCCAGTGGATTGCTTCCGTTGAAGT 512
    |||||
DB 555 TTACGGTAACATGATGAGAAATACGGTCTCACCGGTAGATAATATCCATCATGCAAGAAAT 614
    |||||
QY 513 TGTGG 517
    |||||
DB 615 GGTGG 619
```

```

US-09-938-842A-1762
; Sequence 1762, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1762
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1762

Query Match
Best Local Similarity 50.0%; Pred. No. 0.54; Length 1599;
Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 297 GGTGAATCGGTAGACAGATAGAGTTACTGCTCTCTCGGTGACCAACTG 356
    |||||
DB 420 GGTGACGTTGATTGACAGTACAGTACAGTGTGCGTCACTGCTTACCATCGGAGA 479
    |||||
```


Mon Jun 16 09:24:52 2003

us-09-998-284-1.rmpb

Page 9

Search completed: June 15, 2003, 06:21:33
Job time : 260 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 15, 2003, 01:27:17 ; Search time 76 Seconds
(without alignments)
6633.903 Million cell updates/sec

Title: US-09-998-284-1

Sequence: 1 atgcgtactctgcacacaaa.....ctaacgacgataatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfillseq.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1279.2	77.8	1801	4	US-08-669-304-30 Sequence 30, Appl
2	60	3.6	80161	3	US-09-036-987A-1 Sequence 1, Appl
3	60	3.6	80161	4	US-09-370-700-1 Sequence 1, Appl
4	36.4	2.2	3680	4	US-09-647-390-15 Sequence 15, Appl
5	35.8	2.2	7766	4	US-09-125-619-3 Sequence 3, Appl
6	35	2.1	1632	1	US-08-155-906-1 Sequence 1, Appl
7	33.8	2.1	1981	4	US-09-647-390-17 Sequence 17, Appl
8	33.4	2.0	636	1	US-07-836-642-1 Sequence 1, Appl
9	33.4	2.0	636	1	US-08-169-563-1 Sequence 1, Appl
10	33.4	2.0	636	1	US-08-403-379A-5 Sequence 5, Appl
11	33.4	2.0	636	2	US-08-929-414-5 Sequence 5, Appl
12	33.4	2.0	1553	4	US-09-217-490-1 Sequence 1, Appl
13	33.2	2.0	1458	4	US-09-480-921B-3 Sequence 3, Appl
14	33.2	2.0	7218	1	US-08-232-463-14 Sequence 14, Appl
15	33	2.0	490	2	US-08-630-822A-92 Sequence 92, Appl
16	33	2.0	490	2	US-09-005-069-92 Sequence 92, Appl
17	33	2.0	490	4	US-09-171-156A-41 Sequence 41, Appl
18	32.6	2.0	1800	3	US-09-039-773A-1 Sequence 1, Appl
19	32.2	2.0	6585	3	US-08-746-111-4 Sequence 4, Appl
20	32	1.9	1749	4	US-08-481-190-17 Sequence 17, Appl
21	32	1.9	1749	5	PCT-US93-00869-17 Sequence 17, Appl
22	32	1.9	1764	4	US-08-481-190-14 Sequence 14, Appl
23	32	1.9	1764	5	PCT-US93-00869-14 Sequence 14, Appl
24	32	1.9	2389	4	US-08-974-546-2 Sequence 2, Appl
25	32	1.9	4597	4	US-09-221-017B-1093 Sequence 1093, Ap
26	31.8	1.9	1206	1	US-08-318-947A-1 Sequence 1, Appl
27	31.8	1.9	1206	2	US-08-795-303-1 Sequence 1, Appl

28	31.8	1.9	1303	1	US-08-844-154-1 Sequence 1, Appl
29	31.8	1.9	1303	3	US-09-126-192A-1 Sequence 1, Appl
30	31.8	1.9	1344	2	US-08-902-585-1 Sequence 1, Appl
31	31.8	1.9	6156	4	US-08-891-640-1 Sequence 1, Appl
32	31.8	1.9	6156	4	US-09-723-535-3 Sequence 3, Appl
33	31.8	1.9	4403765	4	US-09-103-840A-2 Sequence 2, Appl
34	31.8	1.9	4411529	4	US-09-103-840A-1 Sequence 1, Appl
35	31.4	1.9	703	4	US-08-858-207A-21 Sequence 21, Appl
36	31.4	1.9	1546	3	US-08-961-083-21 Sequence 21, Appl
37	31.4	1.9	6854	4	US-08-961-527-200 Sequence 200, App
38	31.4	1.9	7577	4	US-08-961-527-46 Sequence 46, Appl
39	31.4	1.9	4403765	4	US-09-103-840A-2 Sequence 2, Appl
40	31.4	1.9	4411529	4	US-09-103-840A-1 Sequence 1, Appl
41	31.2	1.9	915	5	PCT-US94-01149-56 Sequence 56, Appl
42	31.2	1.9	2424	4	US-09-234-393-1 Sequence 1, Appl
43	31.2	1.9	2424	4	US-09-360-545-15 Sequence 15, Appl
44	31.2	1.9	2424	4	US-09-865-171-1 Sequence 1, Appl
45	31.2	1.9	2424	4	US-09-398-395A-45 Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-08-669-304-30
Sequence 30, Application US/08669304
Patent No. 6251626
GENERAL INFORMATION:
APPLICANT: Peter Stougaard
APPLICANT: Ole Cai Hansen
TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hunton & Williams
STREET: 1900 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006-1109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/669,304
FILING DATE: 12 July 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,910
FILING DATE: 7 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Stanislaus Aksman
REGISTRATION NUMBER: 28,562
REFERENCE/DOCKET NUMBER: 320,000003
TELEPHONE: (202) 955-1926
TELEFAX: (202) 778-2201
TELEX: No. 62516266
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURES:
NAME/KEY: CDS
LOCATION: 84..1721
US-08-669-304-30
Query Match 77.8%; Score 1279.2; DB 4; Length 1801;

FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 3.6%; Score 60; DB 3; Length 80161;
Best Local Similarity 52.3%; Pred. No. 4.5e-08;
Matches 157; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

QY 341 CCGGTGACACCACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCAAGGTAAGTTTGC 400
DB 16033 CCGGGCGCCGCTCCGGGCGAGGTGTACCGAACCTGTTCAGAACTGGGGCGTGCATCC 15974
QY 401 CAGGTGTTCTCTTACTCCGTCCGTTGGGTGTGCATTTGCGGTGAGAGTGAAGGTA 460
DB 15973 CCAACGGCGCATGTCCCGGGGTGGGCGGACCGGCGACATCTCCGGCGGGGATACGCC 15914
QY 461 TTTTGGCCAGATTGACACGGTTTCCAGTCGATTGTTATCCGTTGTTGAAGTTGCTTA 520
DB 15913 CGCTGTCGGCGGATTCGGTTGGGTGTGTGACTACTTCAAGGCGTGAAGTGTGT-- 15856
QY 521 ACCGACTTTGACCGAAGACTCTGTTCTTAAGTACTTCAAGAGTTCCGAAGTAAAG 580
DB 15855 -GGTCGACCAAGCGCGGTGAAGTGACATCGTCGAGGCCACCGGAATCCACGGGCGCG 15797
QY 581 ACGGTGAGTTGTTTGGGTCACACTGTTGAGGTGAGGTAACTCCGATTATACCA 640
DB 15796 GTCAAGACTTGTGTGGGCGCACCGGTGGCGGCACTTCGGGATCGTACCA 15737

RESULT 3

US-09-370-700-1/C
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H
APPLICANT: Crawford, Mary C
APPLICANT: Madhuri, Krishnamurthy
APPLICANT: Treadway, Patli J
APPLICANT: Turner, Jan R
APPLICANT: Walden, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 3.6%; Score 60; DB 4; Length 80161;
Best Local Similarity 52.3%; Pred. No. 4.5e-08;
Matches 157; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

QY 341 CCGGTGACACCACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCAAGGTAAGTTTGC 400
DB 16033 CCGGGCGCCGCTCCGGGCGAGGTGTACCGAACCTGTTCAGAACTGGGGCGTGCATCC 15974
QY 401 CAGGTGTTCTCTTACTCCGTCCGTTGGGTGTGCATTTGCGGTGAGAGTGAAGGTA 460
DB 15973 CCAACGGCGCATGTCCCGGGGTGGGCGGACCGGCGACATCTCCGGCGGGGATACGCC 15914
QY 461 TTTTGGCCAGATTGACACGGTTTCCAGTCGATTGTTATCCGTTGTTGAAGTTGCTTA 520
DB 15913 CGCTGTCGGCGGATTCGGTTGGGTGTGTGACTACTTCAAGGCGTGAAGTGTGT-- 15856
QY 521 ACCGACTTTGACCGAAGACTCTGTTCTTAAGTACTTCAAGAGTTCCGAAGTAAAG 580
DB 15855 -GGTCGACCAAGCGCGGTGAAGTGACATCGTCGAGGCCACCGGAATCCACGGGCGCG 15797
QY 581 ACGGTGAGTTGTTTGGGTCACACTGTTGAGGTGAGGTAACTTCGGTATATACCA 640
DB 15796 GTCAAGACTTGTGTGGGCGCACCGGTGGCGGCACTTCGGGATCGTACCA 15737

RESULT 4

US-09-647-390-15
Sequence 15, Application US/09647390
Patent No. 6465636
GENERAL INFORMATION:
APPLICANT: Stulver, Maarten
APPLICANT: Custer, Jerome
APPLICANT: Simons, Lambertus
TITLE OF INVENTION: Pathogen-Inducible Promoter
FILE REFERENCE: MOG 57707/US
CURRENT APPLICATION NUMBER: US/09/647,390
FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: EP 98201024.1
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: PCT/EP99/02178
PRIOR FILING DATE: 1999-03-25
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 15
LENGTH: 3680
TYPE: DNA
ORGANISM: Helianthus annuus
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1889)
NAME/KEY: CDS
LOCATION: (1890)..(3503)
US-09-647-390-15

Query Match 2.2%; Score 36.4; DB 4; Length 3680;
Best Local Similarity 55.6%; Pred. No. 0.31; 56; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 394 GTTTTCCAGTGTCTCTTACTCCGTGGTTGGGTGTGCATTTGCGTGAGGT 453
DB 2412 GGTTCGCGCAGGGGTTGTCAACGTTGGCGTTGGGCACTTTAGTGTGTGG 2471
QY 454 GACGTAATTTGGCAGATTGACAGGTTTCCAGTCGATTGTTATCCGTTGAAGTT 513
DB 2472 TATGTAATTTGATGAGAAATATGTTGTGGTTGATATATGTTGATGCTCAATA 2531
QY 514 GTGCTT 519
DB 2532 ATAGAT 2537

RESULT 5

US-09-125-619-3
Sequence 3, Application US/09125619
Patent No. 6437116
GENERAL INFORMATION:
APPLICANT: NORRIS, STEVEN J.

APPLICANT: JING-REN, ZHANG
APPLICANT: HARDHAM, JOHN M.
APPLICANT: HOWELL, JERRILYN K.
APPLICANT: BARBOUR, ALAN G.
APPLICANT: WEINSTOCK, GEORGE M.
TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
FILE REFERENCE: UTSH:234
CURRENT APPLICATION NUMBER: US/09/125,619
CURRENT FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 7766
TYPE: DNA
ORGANISM: Borrelia burgdorferi
FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 2.2%; Score 35.8; DB 4; Length 7766;
Best Local Similarity 49.2%; Pred. No. 0.78; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 1190 CTAGTCCGAATGAGAGATGCTCTCTTCTTCAAGTTGATGTTGGTGTGAGATTACA 1249
Db 1883 CTGCTGCTATTTGGAAGGTATCGGATGATGTTGGATTTTGGTATGAGATGAGA 1942
Qy 1250 AGTGTGTTGGAGTCTACTGCACTGCTCTGAGAGAGTACATCATTAACCTGCACTACC 1309
Db 1943 AGGATGATGATGATGCTGCTGCTATGCTTTGAGGGGAGTGGCTAAGGATGGAAGTTTG 2002
Qy 1310 AGACATCTGCGAGAGAGAGAGATGATGATTAATTGAATGATTAAGACTTTT 1369
Db 2003 CTGTGAAGAGAGATGAAGAGAGAGCTAGAGGGGCTATTAAAGGAGTACGAGTTGT 2062
Qy 1370 ACGAGAGATG 1380
Db 2063 TCGATTAAGCTG 2073

RESULT 6
US-08-155-906-1/c
Sequence 1, Application US/08155906
Patent No. 5405777
GENERAL INFORMATION:
APPLICANT: ICHIHARA, SHIGEYUKI
APPLICANT: MIZUNO, TAKESHI
TITLE OF INVENTION: ACETIC ACID ASSIMILATING GENE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/155,906
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/850,909
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5405777man F.

REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-552-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1632 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-155-906-1

Query Match 2.1%; Score 35; DB 1; Length 1632;
Best Local Similarity 49.4%; Pred. No. 0.51; Indels 88; Gaps 0;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 1445 AGGAGTCTACTTCACTACCTCCGATGATGATTAACCACTGGAAGAGGTAAGTATG 1504
Db 1573 AANNNTTCAGCATCACTAATCTCACCCTTATGTTTCAAAATATTAACGAGAAATTAAT 1514
Qy 1505 GTCCCTTGAACCTTACTTTTGGGTAACCTGAACGATTTGATCAAGCCAAATGTTGT 1564
Db 1513 TAACCTGAAAGCTAAGTATGCTTCTCGGCCCAATGATGATGATGATGATGATGATG 1454
Qy 1565 GGGATCTCAACGAGATCTTCAACAACAACAGCTATCCCTTAACCTTA 1618
Db 1453 AGAACGTTAAATAATGTTTAACTCCGCAACAGCATCAACAAGAACCGTTA 1400

RESULT 7

US-09-647-390-17
Sequence 17, Application US/09647390
Patent No. 6465636
GENERAL INFORMATION:
APPLICANT: Stulver, Maarten
APPLICANT: Custers, Jerome
APPLICANT: Simons, Lambertus
TITLE OF INVENTION: Pathogen-Inducible Promoter
FILE REFERENCE: MOG 57707/UST
CURRENT APPLICATION NUMBER: US/09/647,390
CURRENT FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: EP 98201024.1
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: PCT/EP99/02178
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 17
LENGTH: 1981

TYPE: DNA
ORGANISM: Lactuca sativa
FEATURE:
NAME/KEY: CDS
LOCATION: (7) ..(1626)
NAME/KEY: unsure
LOCATION: (372)
OTHER INFORMATION: replace (372, "g")
NAME/KEY: unsure
LOCATION: (379)
OTHER INFORMATION: replace (379, "g")
NAME/KEY: unsure
LOCATION: (786)
OTHER INFORMATION: replace (786, "t")
NAME/KEY: unsure
LOCATION: (1105) ..(1106)
OTHER INFORMATION: replace (1105..1106, "ga" or "gg" or "aa")
US-09-647-390-17

Query Match 2.1%; Score 33.8; DB 4; Length 1981;
Best Local Similarity 52.5%; Pred. No. 1.4;

	Matches	74;	Conservative	0;	Mismatches	67;	Indels	0;	Gaps	0;
QY	379	AGAGACCA	CGTAGATTTT	CCCAAGTGTTC	TACTCCGTCGTTGGTGTCAC	438				
Db	511	AAAGCAAC	AGTCATGCTTTT	CCGGCTGGCGTTT	CCCTACTGTTGGAGTTGGTGGCAT	570				
QY	439	ATTGTCG	TGGAGGTGACG	TAATTTTGGCCAGAT	TTCACGTTTGGCAGTGGATTGGTTA	498				
Db	571	TTTAGTGT	GTGTGTATTAG	GTAACTGTATGG	AAATAACGGCCCTTCTGTGTGACAATATT	630				
QY	499	TCCGTTG	TGAAGTTTGCTT	519						
Db	631	GTCGATG	CTCAGTTAATGCAT	651						

RESULT 8
US-07-836-642-1/c
; Sequence 1, Application US/07836642

Patent No. 5304371
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
TITLE OF INVENTION: Against T. cruzi Infection
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University street
CITY: Seattle
STATE: WA
COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/836,642
FILING DATE: 14-FEB-1992

ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: RE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
SPANNEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: *Typanosoma cruzi*
IMMEDIATE SOURCE:

```

; CLONE: Tcd
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..628
;
US-07-836-642-1

```

Query Match	2.0%	Score	33.4	DB	1	Length	636
Best Local	Similarly	Pred. No.	0.93				
Matches	79	Conservative	0	Mismatches	76	Indels	0
						Gaps	0

208 GGAACCGTACAGATCGTTTCTGTGTGTAACGTTCAGAAAGCTTCGTTTTCAGACGAATGT 267
 555 GGGTCGCGGATTTCCGCTCCGCTGGTTTGGGCTCCCGTACCTCGAGTCCGCGGGTTTA 456
 268 GTCAAGGCTATTATCAACGTTACTGGTTGGTTGATCTGGTTTACGACGACGATAGAGGT 327

Db 495 GGGCCCGCGATTTGACCTCCGCGTGGTGGGCTGTGCTGATTTGGGCTCCGCTGGTTTG 436
 QY 328 TACTTGTCTCTTCGCGTACACCACTGGGGTTT 362
 Db 435 GGCTCCGCGTGAATTTCCGCTCCGCGTGGTTTGGGCT 401

RESULT 9
US-08-169-563-1/c
; Sequence 1, Application US/08169563

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1 STATE: WA
2 COUNTRY: USA
3 ZIP: 98101
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: Apple Macintosh
8 OPERATING SYSTEM: Apple Operating System 7.1
9 SOFTWARE: Microsoft Word For Apple, Version 5.1a
10
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/169,563
13 FILING DATE: 17-DEC-1993

```

; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/836,642
; FILING DATE: 14-FEB-1992
;

```

1 CLASSIFICATION: 435
2 ATTORNEY/AGENT INFORMATION:
3 NAME: Perkins, Patricia A
4 REGISTRATION NUMBER: 34,653
5 REFERENCE/DOCKET NUMBER: REED-A
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (206)557-0430
8 INFORMATION FOR SEQ ID NO: 1:

```

? SEQUENCE CHARACTERISTICS:
? LENGTH: 636 base pairs
? TYPE: nucleic acid
? STANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Trypanosoma cruzi
? IMMEDIATE SOURCE:
? CLONE: TCD
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 8..628
?
? JS-08-169-563-1

```

Query Match	2.0%;	Score 33.4;	DB 1;	Length 636;
Best Local Similarity	51.0%;	Pred. No. 0.93;		
Matches 79;	Conservative	0;	Mismatches	76;
			Indels	

QY	208	GGTACCGGTGAGAAATCGTTTCTGGGGTGCACGTGTAACGAAGCTTGTTTTCGACGAATGT	267
Dh	555	GGCTCCGCTGATTTTCGGCTCCGCGGGTTTGGGCTCCGCTCGCTTCGCTCCGCTGATTTA	496
QY	268	GTCAAGCTATTATCAACGTTACTCGTTTGTTGATCTGGTTACGACGACGATAGAGT	327
Dh	495	GGCCCGCGTATTTTCGACTCCGCGGGTTTGGAGCTCTGTGTGATTTTCGGCTCCGCTGATTTG	436

QY 328 TACTTCGTCCTTCCTCCGGTGAACCACTGGGGGTTTC 362
Db 435 GGCTCCGCTGATTTCCGCTCCGCTGGTTGGGGCTC 401

RESULT 10

US-08-403-379A-5/c
; Sequence 5, Application US/08403379A
; Patent No. 5756662
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,379A
; FILING DATE: 14-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..628
; US-08-403-379A-5

Query Match 2.0%; Score 33.4; DB 1; Length 636;
Best Local Similarity 51.0%; Pred. No. 0.93;
Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 208 GGTACCGTCAGATCGTTTCTGGTGCACCTGTTACGAGAAGCTCGTTTTCAGCAATGT 267
Db 555 GGCTCCGCTGATTTCCGCTCCGCTGGTTTGGCTCCGCTGCTCCGCTGGGTTTA 496
QY 268 GTCAAGGCTATTATCAACGTTACTGTTGGTTGAATCTGTTACGACGAGATAGAGT 327
Db 495 GGCCCGCTGATTTCCGCTCCGCTGGTTTGGGCTCTGCTGATTTCCGCTCCGCTGGTTG 436
QY 328 TACTTCGTCCTTCCTCCGGTGAACCACTGGGGGTTTC 362
Db 435 GGCTCCGCTGATTTCCGCTCCGCTGGTTTGGGGCTC 401

RESULT 11
US-08-929-414-5/c
; Sequence 5, Application US/08929414
; Patent No. 5942403
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Haughton, Raymond
; APPLICANT: Skeiky, Yaelir A.W.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
; OF T. CRUZI INFECTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,414
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.406C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8..628
; US-08-929-414-5

Query Match 2.0%; Score 33.4; DB 2; Length 636;
Best Local Similarity 51.0%; Pred. No. 0.93;
Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 208 GGTACCGTCAGATCGTTTCTGGTGCACCTGTTACGAGAAGCTCGTTTTCAGCAATGT 267
Db 555 GGCTCCGCTGATTTCCGCTCCGCTGGTTTGGCTCCGCTGATTTCCGCTGGGTTTA 496
QY 268 GTCAAGGCTATTATCAACGTTACTGTTGGTTGAATCTGTTACGACGAGATAGAGT 327
Db 495 GGCCCGCTGATTTCCGCTCCGCTGGTTTGGGCTCTGCTGATTTCCGCTCCGCTGGTTG 436
QY 328 TACTTCGTCCTTCCTCCGGTGAACCACTGGGGGTTTC 362
Db 435 GGCTCCGCTGATTTCCGCTCCGCTGGTTTGGGGCTC 401

RESULT 12

US-09-217-490-1
; Sequence 1, Application US/09217490
; Patent No. 6165761
; GENERAL INFORMATION:
; APPLICANT: Schneider, Palle
; APPLICANT: Christensen, Soren
; APPLICANT: Dybdal, Lone
; APPLICANT: Fuglsang, Claus Crone
; APPLICANT: Xu, Feng
; APPLICANT: Golightly, Elizabeth
; TITLE OF INVENTION: Carbohydrate Oxidase And use Thereof in
; FILE REFERENCE: 5421.200-US
; CURRENT APPLICATION NUMBER: US/09/217,490
; EARLIER FILING DATE: 1998-12-21
; EARLIER APPLICATION NUMBER: PA 1997 01505
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: PA 1998 00763

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 02:43:57 ; Search time 2460 Seconds
(without alignments)
10823.325 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atcgctactcttcgcacacaa.....ctaagcagactaataatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oher:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	42.8	2.6	646	17	BH695306	BH695306 BOMLJ48TR
2	42.8	2.6	709	17	BH667535	BH667535 BOHZ13TR
3	42.8	2.6	847	17	BH432838	BH432838 BOGVZ24TF
4	42.6	2.6	420	17	BH813311	BH813311 SALK_0639
5	42.6	2.6	447	10	AV786257	AV786257 AV786257
6	42.6	2.6	640	10	AV784340	AV784340 AV784340

Result No.	Score	Match	Length	DB	ID	Description
7	42	2.6	400	10	AV811765	AV811765 AV811765
8	42	2.6	440	10	AV798160	AV798160 AV798160
9	42	2.6	442	10	AV798159	AV798159 AV798159
10	41.8	2.5	748	14	BO841497	BO841497 WHEA211 F
11	41.4	2.5	409	14	BO60219	BO60219 Tsl168 Th
12	41.4	2.5	1101	17	CNS00LOO	AL068607 Drosoph11
13	41	2.5	406	10	AV805562	AV805562 AV805562
14	41	2.5	416	12	BF202701	BF202701 WHE177 C
15	41	2.5	569	14	BU013165	BU013165 OGJ3015 Y
16	41	2.5	595	14	BU007777	BU007777 GGH507 Y
17	41	2.5	688	14	BO990898	BO990898 OGF21G13
18	41	2.5	807	17	AQ957642	AQ957642 LERA085TF
19	40.4	2.5	360	10	AV440363	AV440363 AV440363
20	40.2	2.4	360	13	BI594749	BI594749 As Cgz 43
21	40.2	2.4	360	14	D76262	D76262 CEUK116BF
22	40.2	2.4	387	12	BF428817	BF428817 WHE1710 H
23	40.2	2.4	493	13	BM170559	BM170559 EST573082
24	40.2	2.4	537	13	BM277996	BM277996 As Cgz 51
25	40.2	2.4	600	12	BF485368	BF485368 WHE2310 G
26	40.2	2.4	633	14	BO607036	BO607036 BRY 2917
27	40.2	2.4	769	10	BE414425	BE414425 SCU009 F0
28	40	2.4	350	13	BI594738	BI594738 As Cgz 43
29	40	2.4	492	9	AI995621	AI995621 701675986
30	40	2.4	524	13	BI594743	BI594743 As Cgz 43
31	40	2.4	565	12	BF049993	BF049993 As Cgz 20
32	39.6	2.4	647	9	AL509182	AL509182 AL509182
33	39.6	2.4	770	17	BH478648	BH478648 BOHR61TR
34	39.4	2.4	360	9	AJ470390	AJ470390 AJ470390
35	39.4	2.4	437	17	BH611048	BH611048 SALK_0297
36	39.4	2.4	540	9	AJ470389	AJ470389 AJ470389
37	39.4	2.4	542	14	BO468825	BO468825 HM02H09C
38	39.4	2.4	573	13	BJ471443	BJ471443 BJ471443
39	39.4	2.4	577	14	BO839404	BO839404 WHE165 G
40	39.4	2.4	603	13	BJ479961	BJ479961 BJ479961
41	39.4	2.4	608	17	BO5851	BO5851 GSRL-72a8-u
42	39.4	2.4	624	13	BM441131	BM441131 EBed02 SQ
43	39.4	2.4	632	17	BH605523	BH605523 BOHAE53TR
44	39.4	2.4	634	10	AV938215	AV938215 AV938215
45	39.4	2.4	649	10	AV941742	AV941742 AV941742

ALIGNMENTS

RESULT 1
BH695306/c
LOCUS BOMLJ48TR BO_2_3_KB Brassica oleracea genomic clone BOMLJ48, DNA
DEFINITION
ACCESSION BH695306
VERSION BH695306.1 GI:18767922
KEYWORDS
SOURCE
ORGANISM
GSS.
Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 646)
AUTHORS Town,C.D., Van Aken,S., Uterback,T. and Fraese,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOMLJ48TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
1..646
source

/organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOMLJ48"
 /clone_1id="BO_2.3_KB"
 /note="Vector: pHOSt; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOSt using BstXI linkers"

BASE COUNT 123 a 162 c 220 g 141 t

Query Match 2.6%; Score 42.8; DB 17; Length 646;
 Best Local Similarity 50.5%; Pred. No. 0.16;
 Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 312 CGACGAGATAGAGTACTTCTGCTCCGGTGAACCAACAGGGGTTCTTCAAGAC 371
 DB 383 CGTGGCGAGTCCCTTGGGTGGAGCCGGCCACCTCGGCGAGTTTACTACAGAT 324
 QY 372 CTGTTCAGAGACCAAGGTTTTCAGAGTGTTCCTTACTCCGTGGTTGGG 431
 DB 323 ATGGGAAAAAACAAGTCCACGGGTTCCCGCGAGTCTGCCACCGTGGAGCGG 264
 QY 432 TGGTCAATTTGCGTGGAGTACGGTATTTTGGCCAGATTGACGGTTTCCAGTCA 491
 DB 263 AGGTCAATCAGCGCGCGGTGCGTAACTATGATGAGAAAGTACGACTCTCCGTTGA 204
 QY 492 TTGGTTATCCGGTGTGAAGTTGCG 517
 DB 203 CTACGTACCAAGCCCTAAGATCGTTG 178

RESULT 2
 BH667535/c 709 bp DNA linear GSS 19-FEB-2002
 LOCUS BOHZJ13TR BO_2.3_KB Brassica oleracea genomic clone BOHZJ13, DNA
 DEFINITION
 BOHZJ13TR BO_2.3_KB Brassica oleracea genomic clone BOHZJ13, DNA
 sequence.
 ACCESSION BH667535
 VERSION BH667535.1 GI:18726845
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 709)
 Town, C.D., Van Aken, S., Uterback, T. and Frazer, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
 Location/Qualifiers

1..709
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOHZJ13"
 /clone_1id="BO_2.3_KB"
 /note="Vector: pHOSt; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOSt using BstXI linkers"

BASE COUNT 162 a 191 c 151 g 205 t

Query Match 2.6%; Score 42.8; DB 17; Length 709;
 Best Local Similarity 50.0%; Pred. No. 0.17;
 Matches 135; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 1109 CTGCTTACATGATCAAGACTTCCAGACTTCAGATTGATTTATCTGGAATACCTTA 1168
 DB 331 CTGGAATTTCTTCAAGAGCAAGTCAAGTCAAGAAACCAATCCCAAGGAAGTA 272
 QY 1169 CTGAGTTCTTCAAGCGTTTGAATGTCGGAATGAAGATGCTTTCTTCAAGTTGATA 1228
 DB 271 TGAAGAAGCTTTGGGGGAGCAATGTTGAAATTCAGAAACGTTGTGATGCACTGGAAAC 212
 QY 1229 TGTTCGGTGTGATTCACAAAGTGTGTTGGATGCTACTGAGTTGCTCA---GAGAG 1285
 DB 211 CTTACGTTGCGGTGATGATGATAGATTCCGGCCACCGCCAGGCTTTCTTCAACCGGAAG 152
 QY 1286 AGTACATCATCAAACTGCACTACAGACTACTGCGAGGAAGAAAGCAAGATGACCTTA 1345
 DB 151 GAAACTTTGTTCAAGATTCACTTACTTACTAATGTTGAGCGCAACACACAGAAAGCTA 92
 QY 1346 ACTTGAAGTGAATTAAGACTTTTACGAG 1375
 DB 91 GCTTAAATAATGATGAGGAGTTTACGAG 62

RESULT 3
 BH432838 847 bp DNA linear GSS 12-DEC-2001
 LOCUS BOGV224TR BOGV Brassica oleracea genomic clone BOGV224, DNA
 DEFINITION
 BOGV224TR BOGV Brassica oleracea genomic clone BOGV224, DNA
 sequence.
 ACCESSION BH432838
 VERSION BH432838.1 GI:17618559
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 847)
 Town, C.D., Van Aken, S., Uterback, T. and Frazer, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other GSS: BOGV224TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.

FEATURES
 Location/Qualifiers

1..847
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGV224"
 /clone_1id="BOGV"
 /note="Vector: pHOSt; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOSt using BstXI linkers"

BASE COUNT 201 a 236 c 176 g 234 t

Query Match 2.6%; Score 42.8; DB 17; Length 847;
 Best Local Similarity 50.0%; Pred. No. 0.2;
 Matches 135; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 1109 CTGCTTACATGATCAAGACTTCCAGACTTCAGATTGATTTATCTGGAATACCTTA 1168
 DB 305 CTGAGATTCTTCAAGAGCAAGTCAAGTCAAGTCAAGAAACCAATCCCAAGGAAGTA 246
 QY 1169 CTGAGTTCTTCAAGCGTTTGAATGTCGGAATGAAGATGCTTTCTTCAAGTTGATA 1228
 DB 245 TGAAGAAGCTTTGGGGGAGCAATGTTGAATTCAGAAACGTTGTGATGCAAGTGAAC 186
 QY 1229 TGTTCGGTGTGATTCACAAAGTGTGTTGGATGCTACTGCAAGTTGCTCA---GAGAG 1285

Db		185	CTTACGGTGCGTGTATGATAGCATTCGGGAGACC GCCACGGCGTTTCTCACC GGAAAG	126
QY		1286	AGTACATCATCAACTGCAGTACCAAGACATCTGGCAGGAAGAACAAGANTCCAGTTA	1345
Db		125	GAACTTGTTCAAGATTCACTACTTAATCATGTTGGACGAAACACACAGAGCTA	66
QY		1346	ACTTGAAGTGGATTNAGACTTTTACGAGG	1375
Db		65	GCTTAAAAATGATAGGAGGATTTCAGAGG	36
<hr/>				
RESULT 4				
LOCUS	BH813311/c			
DEFINITION	SALK_063965 Arabidopsis thaliana TDNA insertion lines Arabidopsis			
ACCESSION	BH813311			
VERSION	BH813311.1			
KEYWORDS	GSS.			
ORGANISM	GSS.			
SOURCE	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi- s (bases 1 to 420) Alonso,J.M., Lejse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shim,P., , Zimmerman,J. and Ecker,J.R. A Sequence-indexed library of Insertion Mutations in the Arabidopsis Genome Unpublished (2001) Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (Signal) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel.: 858 453 4100 x1752 Fax: 858 558 6379 Email: eckers@salk.edu This is single pass sequence recovered from the left border of TDNA.			
TITLE	Arabidopsis Genome			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (Signal) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel.: 858 453 4100 x1752 Fax: 858 558 6379 Email: eckers@salk.edu This is single pass sequence recovered from the left border of TDNA.			
<hr/>				
FEATURES				
source	Location/Qualifiers			
	1..420			
	/organism="Arabidopsis thaliana"			
	/strain="Columbia"			
	/db_xref="taxon:3702"			
	/clone="SALK_063965"			
	/clone_id="Arabidopsis thaliana TDNA insertion lines"			
	/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"			
BASE COUNT	147 a 78 c 129 t			
ORIGIN				
	Query Match 2.6%; Score 42.6; DB 17; Length 420;			
	Best Local Similarity 62.9%; Pred.No.0.13;			
	Matches 66; Conservative 0; Mismatches 39; Indels 0; Gaps 0			
QY	1519 TACTTTTGGGTAACTGAACGATGTGATCAAGGCCAATGGTTGTGGATCCTAACAG	1578		
Db	408 TATTTCTGGGGAATTTGAAGATGTGATGTAAAGCTAATAATGATCTGATAT	349		
QY	1579 ATCTTCACAAACAACAGCTATCTCCCTAAACCTCTTAAGAG	1623		
Db	348 TTCTTCAAAAAGAACAGAGCATCTCTCGGTTTGGTAAATGTAG	304		
<hr/>				
RESULT 5				
A786257/c				

LOCUS	AV786257	447 bp	mRNA	linear	EST 28-MAR-2002
DEFINITION	AV786257 RAFL6 Arabidopsis thaliana cDNA clone RAFL06-70-H12 3', mRNA sequence.				
ACCESSION	AV786257				
VERSION	AV786257.1	GI:19805047			
KEYWORDS	EST.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 447)				
ATTNORS	Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,Y., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shimagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.				
TITLE	Large scale analysis of Arabidopsis full-length cDNA (2002b)				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Motoaki Seki Plant Functional Genomics Research Group RIKEN Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel: 81-298-36-4359 Fax: 81-298-36-9060 Email: msek@rc.riken.go.jp An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). This clone is in a modified pBluescript vector as a SecI/XhoI insert. Please visit our web site (http://www.gsc.riken.go.jp/e/planf/index_e.html) for further details.				
FEATURES	Location/Qualifiers				
SOURCE	1..447 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="RAFL06-70-H12" /clone_1lb="RAFL6" /dev_stage="plants at various developmental stages from germination to mature seeds" /lab_host="DH10B" /note="Site 1: SacI; Site 2: XhoI, subjected to denaturation (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"				
BASE COUNT	150 a 89 c 69 g 139 t				
ORIGIN					
Query Match	2.6%; Score 42.6; DB 10; Length 447;				
Best Local Similarity	62.9%; Pred. No. 0.14; Mismatches 39; Indels 0; Gaps 0;				
Matches	66; Conservative 0; Mismatches 39; Indels 0; Gaps 0;				
QY	1519 TACTTTTGGGTAACCTGAACAGATGATGATCAAGGCCAAATGGTTGGGATCTTAACGAG 1578				
Db	252 TATTTCTTGGGAAATTGAAGAGATGATGATGATTAAGCTAGATGATGATCTGATAT 193				
QY	1579 ATCTTCACAAACAACAGCTATCTCTCCCTTAACCTCTTAAGAG 1623				
Db	192 TTCTTCAAAAAACGAACAGAGCATTCCTCCGGTGTGTAATGATAG 148				
RESULT 6					
LOCUS	AV784340	640 bp	mRNA	linear	EST 28-MAR-2002
DEFINITION	AV784340 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-18-J21 3', mRNA sequence.				
ACCESSION	AV784340				
VERSION	AV784340.1	GI:19803130			
KEYWORDS	EST.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 640)				
ATTNORS	Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,				

Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). This clone is in a modified pBluescript vector as a SrfI/XhoI insert. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source
Location/Qualifiers
1..640
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAF105-18-021"
/clone_id="RAF15"
/dev_stage="rosette plants"
/lab_host="SOLR"
/note="Site 1: SrfI; Site 2: XhoI; subjected to dehydration-treated (1.2, 5, 10, 24 hr)"
BASE COUNT
190 a 142 c 106 g 202 t
ORIGIN

Query Match
Best Local Similarity 62.6%; Score 42.6; DB 10; Length 640;
Matches 66; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1519 TACTTTGGGTACTGTAACAGATTGATCAAGCCAAATGTTGGATCTTAAGCAG 1578
DB 211 TATTTCTGGGGAATTTGAAGAGATTGATGATGTTAAAGCTAAGTATGATCTCGATMAT 152
QY 1579 ATCTTCACAAACAGAGCTATCTTAACTTAAGCTTAAGAG 1623
DB 151 TTCTTCAAAAACGACAGAGATTCTCCGGTTCGTAAATG 107

RESULT 7
AV811765/c 400 bp mRNA linear EST 29-MAR-2002
LOCUS
DEFINITION
AV811765 RAF19 Arabidopsis thaliana cDNA clone RAF109-69-123 3',
mRNA sequence.
ACCESSION
AV811765
VERSION
AV811765.1 GI:19845750
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 400)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLIC-1 vector (Carninci et

al., submitted for publication) digested with BamHI and Sali. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source
Location/Qualifiers
1..400
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAF109-69-123"
/clone_id="RAF19"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/note="Site 1: BamHI; Site 2: Sali; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
BASE COUNT
115 a 90 c 59 g 136 t
ORIGIN

Query Match
Best Local Similarity 66.7%; Score 42; DB 10; Length 400;
Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1519 TACTTTGGGTACTGTAACAGATTGATCAAGCCAAATGTTGGATCTTAAGCAG 1578
DB 237 TATTTCTGGGGAATTTGAAGAGATTGATGATGTTAAAGCTAAGTATGATCTCGAGAA 178
QY 1579 ATCTTCACAAACAGAGCTATCTTAACTTAAGCTTAAGAG 1608
DB 177 TTCTTCAAAAACGACAGAGATTCTCCCT 148

RESULT 8
AV798160/c 440 bp mRNA linear EST 29-MAR-2002
LOCUS
DEFINITION
AV798160 RAF19 Arabidopsis thaliana cDNA clone RAF109-15-D04 3',
mRNA sequence.
ACCESSION
AV798160
VERSION
AV798160.1 GI:19832143
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 440)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and Sali. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source
Location/Qualifiers
1..440
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAF109-15-D04"
/clone_id="RAF19"
/dev_stage="plants at various developmental stages from

QY 1519 TACTTTTGGGTAACTCGAAGATTATCAAGCCAAATGGTGTGGGATCTCAAGAG 15718
 Db 267 TATTCTTGGGGAATTTGAAGAATTATGATGATTTAAAGTAAGATGATCTCGAAT 208
 QY 1579 ATCTTCACAAACAACAGTCTATCCCTACT 1608
 Db 207 TTCTTCAAAAACGACAGAGTATTCCTCCCT 178
 RESULT 10
 BQ841497 748 bp mRNA linear EST 12-AUG-2002
 DEFINITION WHE4211.F11.L20ZS Aegilops speltoides pre-meiotic anther cDNA
 LOCUS library Aegilops speltoides cDNA clone WHE4211.F11.L20, mRNA
 sequence.
 ACCESSION BQ841497 GI:22210906
 VERSION BQ841497.1
 KEYWORDS EST.
 SOURCE Aegilops speltoides.
 ORGANISM Aegilops speltoides
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Aegilops.
 .1 (bases 1 to 748)
 REFERENCE Ahniov, E., Anderson, O.D., Chao, S., Crossman, C., Devorak, J., Lazo
 AUTHORS , G.R., Pham, J., Rausch, C.J., Wilson, C. and Woo, J.
 TITLE The structure and function of the expressed portion of the wheat
 genomes - Pre-meiotic anther cDNA library from Aegilops speltoides
 JOURNAL Unpublished (2002)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer
 FEATURES
 Source Location/Qualifiers
 1..748
 /organism="Aegilops speltoides"
 /cultivar="F2 from 2-12-4-8-1-1-(1) x PI36909-12-811-(1)"
)
 /db_xref="taxon:4573"
 /clone="WHE4211.F11.L20"
 /clone_1b="Aegilops speltoides pre-meiotic anther cDNA
 library"
 /rissue type="Anther"
 /dev_stage="Pre-meiotic anthers"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
 chamber at the University of California, Davis (Ahniov in
 Devorak's lab). Pre-meiotic anthers were harvested, total
 RNA and poly(A) RNA were prepared, a cDNA library was made
 , and the cDNA clones were in vivo excised to give
 phagescript phagemids at the University of California,
 Davis (Ahniov in Devorak's lab). Plasmid DNA preparations
 and DNA sequencing were performed in the OD Anderson lab
 (all other authors)."
 BASE COUNT 216 a 161 c 206 g 164 t 1 others
 ORIGIN
 Query Match 2.5%; Score 41.8; DB 14; Length 748;
 Best Local Similarity 52.6%; Prid. No. 0.37;
 Matches 91; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
 QY 1409 CTAAACCTCAGGTGAGGTGAAGGTCTTTTGGAGGATGCTTCAACTACCTGTG 1468
 Db 200 CTGACGATGAAGTGAAGATGAAGAGAGGATACTGAGGAGGCGCAAGCTTGAGGAAATTG 259

RESULT 13
AV805562 406 bp mRNA linear EST 29-MAR-2002
LOCUS AV805562 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-43-G11 3',
DEFINITION mRNA sequence.
ACCESSION AV805562
VERSION AV805562.1 GI:19839547
KEYWORDS EST.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 406)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@rc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/planf/index_e.html) for further
details.

FEATURES
source
1. .406
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL09-43-G11"
/clone_1ib="RAF19"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/note="Site 1: BamHI; Site 2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT 122 a 99 c 64 g 121 t
ORIGIN

Query Match 2.5%; Score 41; DB 10; Length 406;
Best Local Similarity 61.9%; Pred. No. 0.4; Indels 0; Gaps 0;
Matches 65; Conservative 0; Mismatches 40;

QY 1519 TACTTTTGGGTAACTGAAAGATTCAGAGCCAAATGTTGGGATCCTTAACGAG 1578
DB 204 TATTTCTTGGGGAATTTGAAGATTAATGAGATTAAGCTAAAGTATGATCTGATTAAT 145
QY 1579 ATCTTCAACAACAAGATCTATCCTACTTAACCTCTTGAAGAG 1623
DB 144 TTCTTCAGAACGACGAGATATTCCTCGGTTCTGTTAAAGTGA 100

RESULT 14
BF202701 416 bp mRNA linear EST 06-NOV-2000
LOCUS BF202701
DEFINITION WHE1777 C01 F01ZS wheat pre-anthesis spike cDNA library Triticum
aestivum cDNA clone WHE1777_C01_F01, mRNA sequence.
ACCESSION BF202701
VERSION BF202701.1 GI:11117443
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
P.S., Heis, C.C., Kang, Y., Lazo, G.R., Miller, R., Raush, C.J.,
Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@w.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Strategene SK primer.

FEATURES
source
1. .416
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1777 C01 F01"
/clone_1ib="wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemid in the T7 Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

BASE COUNT 131 a 78 c 125 g 82 t
ORIGIN

Query Match 2.5%; Score 41; DB 12; Length 416;
Best Local Similarity 52.0%; Pred. No. 0.41; Indels 0; Gaps 0;
Matches 92; Conservative 0; Mismatches 85;

QY 1409 CTAACTCAGGTTGAGAGTGTAAAGTGTTTTGGAGGATGTAATCTTAACCTCTG 1468
DB 214 CTGACGATGAAGATGAGATGAGAAAGAGATTAAGAGGCAAGGTTGAGAAATTG 273
QY 1469 ATGTTGACTTGAACAACCTGGAAGAGGTAAGTATGTCCTTGGAACTTTACTTTGG 1528
DB 274 ATGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGATCAAGAGGTTTCTCAGAGT 333
QY 1529 GTAACTGAACGATGATGATCAAGGCCAAATGTTGGGATCCTTAACGAGATCTTA 1585
DB 334 GGAACCTGATCAACAAGCAAGAACCCATCTGATGAGAGAGCCTGAGAGATCACA 390

RESULT 15
BU013165 569 bp mRNA linear EST 22-AUG-2002
LOCUS BU013165
DEFINITION OGJ3015.Yg.ab1 OG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
OGJ3015, mRNA sequence.
ACCESSION BU013165
VERSION BU013165.1 GI:22447560
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; eunasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.

REFERENCE
AUTHORS

1 (bases 1 to 569)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., Van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compgenomics.ucdavis.edu/>
Unpublished (2002)

JOURNAL
COMMENT

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Armundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_Ca_contig4252, see <http://cgpb.ucdavis.edu/>
for details.
Plate: QG53 row: 0 column: 15.

FEATURES
source

Location/Qualifiers
1..569
/organism="Lactuca sativa"
/cultivar="L. serriola"
/db_xref="taxon:4236"
/clone="QG3015"
/clone_id="QG_EFGHU lettuce serriola"
/lab_host="E.coli"
/note="Vector: pBRCDNA5fiAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at <http://cgpb.ucdavis.edu/>
TAG_Lib=QG_EFGHU lettuce serriola
TAG_Tissue=flowers pre-fertilized
TAG_SEQ=GCTTGACGGG"

BASE COUNT
ORIGIN

166 a 81 c 131 g 191 t

Query Match 2.5%; Score 41; DB 14; Length 569;
Best Local Similarity 60.2%; Pred. No. 0.52;

Matches 68; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1302 GCACTTACGACGACTGCGAGAGAGACAGAGATGAGTTACTTGAAGTGATTAG 1361

DB 51 GAAATATCGACACACTTTCTGAATACATTAAGGCTGAGTTGAAGCAGACACATTGA 110

QY 1362 AGACTTTTACGAGAGATGATGAGCCTTATGATGATGTTCCAGAGCCCTAACA 1414

DB 111 AGAGTCTACAGAGAGATGATGAGCCTTATGATGATGATCAACCCCTAAGA 163

Search completed: June 15, 2003, 03:39:35
Job time : 2465 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 00:41:27 ; Search time 388 Seconds

(without alignments)
9541.972 Million cell updates/sec

Title: US-09-998-284-1

Perfect score: 1644
Sequence: 1 atgcgactctgcacacaaa.....ctaaagactaataatagtag 1644

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N.Geneseq.101002.*
2: /SID2/gcgdata/geneq/geneq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneq/geneq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneq/geneq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneq/geneq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneq/geneq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneq/geneq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneq/geneq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneq/geneq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneq/geneq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneq/geneq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneq/geneq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneq/geneq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneq/geneq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneq/geneq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneq/geneq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneq/geneq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneq/geneq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneq/geneq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneq/geneq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneq/geneq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneq/geneq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1644	100.0	1644	22	Hexose oxidase (HO)
2	1644	100.0	1644	22	Chondrus crispus he
3	1632.8	99.3	1644	21	Synthetic hexose o
4	1279.2	77.8	1801	18	DNA encoding hexos
5	126	7.7	126	21	Primer hox16 used
6	126	7.7	126	22	PCR primer 6 used
7	120	7.3	120	21	Primer hox5 used
8	120	7.3	120	21	Primer hox14 used
9	120	7.3	120	22	PCR primer hox5a u

ALIGNMENTS

C	10	120	7.3	120	22	AA506166	PCR primer 4 used
C	11	118	7.2	118	21	AA294019	Primer hox9 used
C	12	118	7.2	118	22	AA506161	PCR primer hox9a u
C	13	117	7.1	117	21	AA294029	Primer hox19 used
C	14	117	7.1	117	22	AA506171	PCR primer 9 used
C	15	116.4	7.1	118	22	AA506162	PCR primer hox10b
C	16	116	7.1	116	22	AA294018	Primer hox8 used
C	17	116	7.1	116	22	AA506160	PCR primer hox8b u
C	18	113.2	6.9	118	21	AA294020	Primer hox10 used
C	19	111	6.8	111	21	AA294028	Primer hox18 used
C	20	111	6.8	111	22	AA506170	PCR primer 8 used
C	21	109	6.6	109	21	AA294016	Primer hox6 used
C	22	109	6.6	109	21	AA294017	Primer hox7 used
C	23	109	6.6	109	22	AA506158	PCR primer hox6b u
C	24	109	6.6	109	22	AA506159	PCR primer hox7a u
C	25	108	6.6	108	21	AA294025	Primer hox15 used
C	26	108	6.6	108	21	AA294027	Primer hox17 used
C	27	108	6.6	108	21	AA294031	Primer BgIII-NsII
C	28	108	6.6	108	22	AA506167	PCR primer 5 used
C	29	108	6.6	108	22	AA506169	PCR primer 7 used
C	30	107	6.5	107	21	AA294012	Primer hox2 used
C	31	107	6.5	107	22	AA506154	PCR primer hox2B u
C	32	106	6.4	106	22	AA506153	PCR primer hox1A u
C	33	106	6.4	106	22	AA506155	PCR primer hox3a u
C	34	106	6.4	106	22	AA506156	PCR primer hox4b u
C	35	104.4	6.4	106	21	AA294014	Primer hox4 used
C	36	102	6.2	102	22	AA506164	PCR primer 2 used
C	37	100	6.1	100	21	AA294032	Primer BgIII-NsII
C	38	98.8	5.7	102	21	AA294022	Primer hox12 used
C	39	94	5.7	105	21	AA294013	Primer hox3 used
C	40	92.8	5.6	96	21	AA294021	Primer hox11 used
C	41	92.8	5.6	96	22	AA506163	PCR primer 1 used
C	42	85.2	5.2	90	21	AA294023	Primer hox13 used
C	43	85.2	5.2	90	22	AA506165	PCR primer 3 used
C	44	72.2	4.4	83	21	AA294037	Primer SKI-sense
C	45	68.2	4.1	83	21	AA294038	Primer SKL-antisense

RESULT 1
AA506173 standard; DNA; 1644 BP.

AC AA506173;

DT 12-SEP-2001 (first entry)

DE Hexose oxidase (HOX) synthetic gene sequence.

KW HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;

KW Food manufacturing; beverage; detergent; baking; dough improving agent;

KW D-hexose:O2-oxidoreductase; ds.

OS Chondrus crispus.

OS Synthetic.

FT Key Location/Qualifiers
CDS 1..1641
/tag= a
/product= "Hexose reductase (HOX)"
/EC_number= "EC 1.1.3.5"

PN WO200138544-A1.

PD 31-MAY-2001.

PF 24-NOV-2000; 2000WO-1B01886.

PR 24-NOV-1999; 99GB-0027801.

PA (DANI-) DANISCO AS.

XX Johanen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
 PI Zargahi MR;
 XX WPI, 2001-367695/38.
 DR P-PSDB; AAU02192.
 XX Releasing soluble or membrane associated intracellular protein from a
 PT cell for manufacturing food, comprises contacting the cell with a
 PT membrane extracting composition and causing the protein to be released
 PT in soluble form -
 XX
 XX Claim 17, Fig 6; 108pp; English.
 XX The sequence represents the coding sequence of synthetic hexose oxidase
 CC (D-hexose:O₂-oxidoreductase, EC 1.1.3.5), also referred to as Hox, gene.
 CC The native Hox gene was altered using site-directed mutagenesis in order
 CC to match the codon usage to known codon preferences of biotechnologically
 CC relevant yeasts, such as *Pichia sp.*, to facilitate high level production
 CC in these organisms. The invention involves a method for releasing a
 CC soluble or membrane associated intracellular protein of interest (POI)
 CC from a cell involving contacting a cell comprising a soluble or membrane
 CC associated intracellular POI with a membrane extracting composition (I)
 CC and causing the POI to be released from the cell in a soluble form. The
 CC method is useful for releasing POI, such as an interleukin I receptor
 CC antagonist (IL-1ra) which involves contacting a transformed cell
 CC comprising IL-1ra with (I) and causing IL-1ra to be released from the
 CC transformed cell, in a soluble form. The method is also useful for
 CC screening mutated cells or transformed cells producing elevated levels of
 CC intracellular POI. The method is used to release a POI for manufacturing
 CC food products, such as beverages, preparation of detergents, and in
 CC baking as a dough improving agent. The method obtains a fast, specific
 CC and economically efficient extraction of a soluble or membrane associated
 CC intracellular POI without the use of conventional cell disruption
 CC techniques. The resulting cell extract contains less contaminating
 CC intracellular DNA and is relatively free of cell wall fragments. The
 CC intracellular POI can be recovered from a eukaryotic host organism such
 CC as yeast, before glycosylation takes place. The method can be used to
 CC prevent contact of intracellular POI with the extracellular growth
 CC medium.
 CC
 XX Sequence 1644 BP; 430 A; 337 C; 404 G; 473 T; 0 other;
 SQ
 Query Match 100.0%; Score 1644; DB 22; Length 1644;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 TCCTCAAGACCTTGTTCAGAGACCAAGGTAGAGTTTCCAGAGTGTCTGTACTCC 420
 Qy 421 GTCCGTTGGGTGTCACATTGTCGGTGAAGGACGATATTGGCCAGATTGACGGT 480
 Db 421 GTCCGTTGGGTGTCACATTGTCGGTGAAGGACGATATTGGCCAGATTGACGGT 480
 Qy 481 TTGCGAGTCGATTTGTTATCCGGTGTGGAAGTTGTCGTTAAGCACTTTGACCGAAGC 540
 Db 481 TTGCGAGTCGATTTGTTATCCGGTGTGGAAGTTGTCGTTAAGCACTTTGACCGAAGC 540
 Qy 541 TCTGTTCTTAAGTACCTTCAACAGATTCGGAAGGTACAGAGGTGATGTTTGGGCT 600
 Db 541 TCTGTTCTTAAGTACCTTCAACAGATTCGGAAGGTGATGATGTTTGGGCT 600
 Qy 601 CACACTGTGAGAGTGAAGTAACTTCGGTATTATCACAATTAATTAATTAAGATTG 660
 Db 601 CACACTGTGAGAGTGAAGTAACTTCGGTATTATCACAATTAATTAATTAAGATTG 660
 Qy 661 CCAATGTCCTCAAGAGGTGTCATCGCTTAACTTAACCTTCTTGGAGCGGTTCACT 720
 Db 661 CCAATGTCCTCAAGAGGTGTCATCGCTTAACTTAACCTTCTTGGAGCGGTTCACT 720
 Qy 721 AGAGATCCCTTGCAAGATTTGTTGACTTAAGTACTTAAGTGGCTGATGTTGAAG 780
 Db 721 AGAGATCCCTTGCAAGATTTGTTGACTTAAGTACTTAAGTGGCTGATGTTGAAG 780
 Qy 781 AATACCTGTTGTAAGTTCACAAATCTTCCACCAAGCAGTGAAGATTGTTATGACTTG 840
 Db 781 AATACCTGTTGTAAGTTCACAAATCTTCCACCAAGCAGTGAAGATTGTTATGACTTG 840
 Qy 841 TATACATCTTACTCTTAACAGCGCCGAGAGAAAGTTGCCAAGACAGACATATCAATTG 900
 Db 841 TATACATCTTACTCTTAACAGCGCCGAGAGAAAGTTGCCAAGACAGACATATCAATTG 900
 Qy 901 GAGCGTCACTTGAACAGATCTTCAAAAACATGCGACCTTACCAAGCTTGTGGTCA 960
 Db 901 GAGCGTCACTTGAACAGATCTTCAAAAACATGCGACCTTACCAAGCTTGTGGTCA 960
 Qy 961 GCTGTTGGGCTCTTCCCTGTTAGACCTAGAAAGACACATCAAGACTTCTTAT 1020
 Db 961 GCTGTTGGGCTCTTCCCTGTTAGACCTAGAAAGACACATCAAGACTTCTTAT 1020
 Qy 1021 ATGCATGACGAGACTATGACATACCTTCTAAGCTTTGACTGAGACTATCAACGGTTCC 1080
 Db 1021 ATGCATGACGAGACTATGACATACCTTCTAAGCTTTGACTGAGACTATCAACGGTTCC 1080
 Qy 1081 GGTCTTAATCAGAGAGTAAATCAAGTCTGCTTAATGATCAAGACTTTCCAGACTTC 1140
 Db 1081 GGTCTTAATCAGAGAGTAAATCAAGTCTGCTTAATGATCAAGACTTTCCAGACTTC 1140
 Qy 1141 CAGATTGATGTTATCTGGAATATCTTACTAGAGTCTGACGGTTGACTAGTGCAGAA 1200
 Db 1141 CAGATTGATGTTATCTGGAATATCTTACTAGAGTCTGACGGTTGACTAGTGCAGAA 1200
 Qy 1201 ATGAGAGATGCTCTTCTTCAAGTTGATGTTGATGTTGATGAGATTCAAGATTGTTGG 1260
 Db 1201 ATGAGAGATGCTCTTCTTCAAGTTGATGTTGATGTTGATGAGATTCAAGATTGTTGG 1260
 Qy 1261 GATGCTACTGCACTTCTGAGAGAGTACATCATCAAACTGCAAGTACAGACATCTGG 1320
 Db 1261 GATGCTACTGCACTTCTGAGAGAGTACATCATCAAACTGCAAGTACAGACATCTGG 1320
 Qy 1321 CAGAGAAAGACAGAGATGAGTAACTTGAAGTGAATTAGACTTTTACAGAGAGATG 1380
 Db 1321 CAGAGAAAGACAGAGATGAGTAACTTGAAGTGAATTAGACTTTTACAGAGAGATG 1380
 Qy 1381 TATGACCTTATGTTGTTGTTTCAAGACCTTAACCTCAGATTGAGAGTGAAGTGT 1440
 Db 1381 TATGACCTTATGTTGTTGTTTCAAGACCTTAACCTCAGATTGAGAGTGAAGTGT 1440
 Qy 1441 TTTGAGGATGCTACTTCAACTACCTGATGTTGACTTGAACAACCTGGAAGCGGTAA 1500
 Db 1441 TTTGAGGATGCTACTTCAACTACCTGATGTTGACTTGAACAACCTGGAAGCGGTAA 1500

QY 1501 TATGTCCTTGGAACTTACTTTTGGTAACTTGAAAGATTCATCAAGCCAAATGG 1560
DB 1501 TATGTCCTTGGAACTTACTTTTGGTAACTTGAAAGATTCATCAAGCCAAATGG 1560
QY 1561 TTGTGGGATCTTAAGAGATCTTCAACAAACAGCTATCTCTAATACTCTTAAG 1620
DB 1561 TTGTGGGATCTTAAGAGATCTTCAACAAACAGCTATCTCTAATACTCTTAAG 1620
QY 1621 GAGCCTAAGCAGACTAATATGATG 1644
DB 1621 GAGCCTAAGCAGACTAATATGATG 1644

RESULT 2

AAF23590
ID AAF23590 standard; DNA; 1644 BP.

AAF23590;

23-MAR-2001 (first entry)

Chondrus crispus hexose oxidase enzyme DNA.

Hexose oxidase; marine alga; anti-fouling; de.

Chondrus crispus.

WO200075293-A2.

14-DEC-2000.

02-JUN-2000; 2000WO-1B00829.

04-JUN-1999; 99GB-0013050.

(DANI-) DANISCO AS.

Poulsen CH, Kragh KM;

WPI; 2001-112148/12.

New anti-fouling composition, useful as a coating for treating different surfaces, e.g. outdoor woodwork, external surface of a central heating system, or a hull of a marine vessel -

Claim 5; Page 35-36; 36pp; English.

The present invention relates to a new anti-fouling composition. The composition involves a surface coating material, a hexose oxidase enzyme obtained from a marine organism and a substrate for the enzyme. The anti-fouling composition is useful as a coating formulated for treating a surface, e.g. outdoor wood work, external surface of a central heating system, or a hull of a marine vessel. It is also useful as an anti-fouling agent for marine structures exposed to seawater flora and fauna.

Sequence 1644 BP; 430 A; 337 C; 404 G; 473 T; 0 other;

Query Match 100.0%; Score 1644; DB 22; Length 1644;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTACTTTGGCCAAAGAGCCAGGTTACATTGTTATGAGCTCAAGCTGTACT 60
DB 1 ATGGCTACTTTGGCCAAAGAGCCAGGTTACATTGTTATGAGCTCAAGCTGTACT 60
QY 61 CCAGACAGCTGAGCCCAAGATTCATCATGAGCAAGGTTTCAACAGAGATGATT 120
DB 61 CCAGACAGCTGAGCCCAAGATTCATCATGAGCAAGGTTTCAACAGAGATGATT 120
QY 121 GGTAACAATCATGATTCGTTTACGTCGTTTACATCCACAAGGTCCTGTACTGCTTGG 180
DB 121 GGTAACAATCATGATTCGTTTACGTCGTTTACATCCACAAGGTCCTGTACTGCTTGG 180

DB 121 GGTAACAATCATGATTCGTTTACGTCGTTTACATCCACAAGGTCCTGTACTGCTTGG 180
QY 181 GACAGAGCTATGAAAGAGTCTTCAGAGTACCGTCAAGATCGTTTGGGTCACTGT 240
DB 181 GACAGAGCTATGAAAGAGTCTTCAGAGTACCGTCAAGATCGTTTGGGTCACTGT 240
QY 241 TACGAAGACTTGTTTGCACGAATGTGTCAAGGCTATTTATCAAGCTTATGCTT 300
DB 241 TACGAAGACTTGTTTGCACGAATGTGTCAAGGCTATTTATCAAGCTTATGCTT 300
QY 301 GAATCTGGTTAGACAGACGATAGAGTTTACCTTCCTTCGAGTACCACTGGGGGT 360
DB 301 GAATCTGGTTAGACAGACGATAGAGTTTACCTTCCTTCGAGTACCACTGGGGGT 360
QY 361 TCCTTCAAGACCTTGTTGAGAGACCAAGGATAGTTTGGCAGGTGCTTCTTACTCC 420
DB 361 TCCTTCAAGACCTTGTTGAGAGACCAAGGATAGTTTGGCAGGTGCTTCTTACTCC 420
QY 421 GTCGGTTTGGGTGTGCATATGTCGAGTGAAGGTGACGATATTTGGCCAGATTGCA 480
DB 421 GTCGGTTTGGGTGTGCATATGTCGAGTGAAGGTGACGATATTTGGCCAGATTGCA 480
QY 481 TTGCAAGTCGATTTGGTTATCCGGTGTGAAGTTGTCTTGAACCCAGTCTTGAAC 540
DB 481 TTGCAAGTCGATTTGGTTATCCGGTGTGAAGTTGTCTTGAACCCAGTCTTGAAC 540
QY 541 TCTGTTCTTAAGTACGTTTCAAGAGATTCGAAAGTACAGAGGTGATGTTTGGGCT 600
DB 541 TCTGTTCTTAAGTACGTTTCAAGAGATTCGAAAGTACAGAGGTGATGTTTGGGCT 600
QY 601 CACACTGTGAGAGTGAAGTAACTTCGATTTATCAACAAATACTACTTCAAGATTTG 660
DB 601 CACACTGTGAGAGTGAAGTAACTTCGATTTATCAACAAATACTACTTCAAGATTTG 660
QY 661 CCAATGTCTCAAGAGGTGTATCGGTTCTTAACCTTCTTTGGGACGTTTCACT 720
DB 661 CCAATGTCTCAAGAGGTGTATCGGTTCTTAACCTTCTTTGGGACGTTTCACT 720
QY 721 AAGAGATGCTTCAAGATTTGTAAGTAACTTCAAGTGGCTAGATGATGTTGGAAG 780
DB 721 AAGAGATGCTTCAAGATTTGTAAGTAACTTCAAGTGGCTAGATGATGTTGGAAG 780
QY 781 AATAGTGTGTAAGTCAAAATCTTCCACCAAGCAGTGAAGATGTTTATGATCTTG 840
DB 781 AATAGTGTGTAAGTCAAAATCTTCCACCAAGCAGTGAAGATGTTTATGATCTTG 840
QY 841 TATACATCTACTCTTAACGAGCCGAGAGAGATTTGCCAAGACAGACATATCATTTG 900
DB 841 TATACATCTACTCTTAACGAGCCGAGAGAGATTTGCCAAGACAGACATATCATTTG 900
QY 901 GAGGCTGAGATTTGAACAGATCTTCAAAACATGCGAGCCCTTCAAGGCTTGTGCTAT 960
DB 901 GAGGCTGAGATTTGAACAGATCTTCAAAACATGCGAGCCCTTCAAGGCTTGTGCTAT 960
QY 961 GCTGTTGGGCTCTTTCCCTGTAGACTTAAAGAGACACATCAAGACTTCTTAT 1020
DB 961 GCTGTTGGGCTCTTTCCCTGTAGACTTAAAGAGACACATCAAGACTTCTTAT 1020
QY 1021 ATGATGACGAGACTATGAGTACCTTTCTAAGCTTTGACTGAGATCTTCAACGCTTCC 1080
DB 1021 ATGATGACGAGACTATGAGTACCTTTCTAAGCTTTGACTGAGATCTTCAACGCTTCC 1080
QY 1081 GGTCTTAATCAAGAGATTAAGTCAAGTGTCTTCAATGATTAAGAGACTTTCCAGACTTC 1140
DB 1081 GGTCTTAATCAAGAGATTAAGTCAAGTGTCTTCAATGATTAAGAGACTTTCCAGACTTC 1140
QY 1141 CAGATTGATTTATCTGAAATACCTTACTAGAGTTCCGACGTTTGAATGATGCGGAA 1200
DB 1141 CAGATTGATTTATCTGAAATACCTTACTAGAGTTCCGACGTTTGAATGATGCGGAA 1200
QY 1201 ATGAAGATGCTCTTCTTCAAGGTTGATGTTGCTGAGATTTCAAGGTTGTTGG 1260
DB 1201 ATGAAGATGCTCTTCTTCAAGGTTGATGTTGCTGAGATTTCAAGGTTGTTGG 1260

QY 1261 GATGCTACTGCAAGTGGCTCAGAGAGATCATCAAACTGACAGTACAGACATCTGG 1320
DB 1261 GATGCTACTGCAAGTGGCTCAGAGAGATCATCAAACTGACAGTACAGACATCTGG 1320
QY 1321 CAGGAGAGACAGAGATGACAGTAACTTGAAGTGAAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 CAGGAGAGAGACAGAGATGACAGTAACTTGAAGTGAAGAGAGAGAGAGAGAGAG 1380
QY 1381 TATGAGCTTATGAGTGGTGTTCAGACCTTAACCTAGGTTGAGAGTGAAGAGTGT 1440
DB 1381 TATGAGCTTATGAGTGGTGTTCAGACCTTAACCTAGGTTGAGAGTGAAGAGTGT 1440
QY 1441 TTTGAGGAGATGCTACTTCAATCACTGATGTTGACTTGAACAAGTGAAGAGAGTAA 1500
DB 1441 TTTGAGGAGATGCTACTTCAATCACTGATGTTGACTTGAACAAGTGAAGAGAGTAA 1500
QY 1501 TATGAGTCCCTTGAAGATCTTCTTTTGGGTAACTTGAAGAGAGAGAGAGAGAG 1560
DB 1501 TATGAGTCCCTTGAAGATCTTCTTTTGGGTAACTTGAAGAGAGAGAGAGAGAG 1560
QY 1561 TTTGAGGATCTTAAAGAGATCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 TTTGAGGATCTTAAAGAGATCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 GAGCCTAAGCAGACTAATAGTAG 1644
DB 1621 GAGCCTAAGCAGACTAATAGTAG 1644

RESULT 3
AAZ94039
ID AAZ94039 standard; DNA; 1644 BP.

AC AAZ94039;

DT 29-AUG-2000 (first entry)

DE Synthetic hexose oxidase gene.

KW Hexose oxidase; production; fermentation; synthetic; modification;

KM prokaryote; eukaryote; ss.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..1641

FT /cag= a

FT /product= Hexose oxidase

XX BP1008651-A2.

XX 14-JUN-2000.

XX 01-DEC-1999; 99EP-0204068.

XX 09-DEC-1998; 98DK-0001630.

XX (BIOT-) BIOTEKNOLOGISK INST.

XX Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;

XX WPI; 2000-389309/34.

XX P-PSDB; AAY83619.

XX Nucleic acid fragment useful for producing large amounts of hexose
XX oxidase comprises nucleotide sequence encoding hexose oxidase that is
XX modified by at least one codon

XX Claim 1; Page 25-27; 42pp; English.

XX Wild type hexose oxidase can be produced by isolating a nucleotide
XX sequence coding for hexose oxidase naturally produced by an organism

CC and modifying the sequence so that the modified sequence is
CC expressed, under identical conditions, at a level at least 10% higher
CC than the non-modified sequence. The method is useful for producing
CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC of producing hexose oxidase were to isolate the enzyme from a source
CC that naturally produces the enzyme, using host organisms which
CC produce a relatively low level of expression. Therefore, industrial
CC production of the enzyme is not feasible using this method. This is a
CC much improved method which produces at least 250 mg enzyme per liter
CC of fermentation medium.

XX Sequence 1644 BP; 430 A; 344 C; 401 G; 469 T; 0 other;

Query Match 99.3%; Score 1632.8; DB 21; Length 1644;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1637; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGCTACTTTGGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
DB 1 ATGGCTACTTTGGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 61 CCAGACAG 120
DB 61 CCAGACAG 120
QY 121 GGTACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 121 GGTACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 181 GACAGAGCTATGAG 240
DB 181 GACAGAGCTATGAG 240
QY 241 TACGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 241 TACGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 301 GAATCTGTTACGACAG 360
DB 301 GAATCTGTTACGACAG 360
QY 361 TCCTTCAAG 420
DB 361 TCCTTCAAG 420
QY 421 GTGAGTTGGAG 480
DB 421 GTGAGTTGGAG 480
QY 481 TTGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
DB 481 TTGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
QY 541 TCTGTTCTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 541 TCTGTTCTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 600
QY 601 CACACTGAG 660
DB 601 CACACTGAG 660
QY 661 CCAATGCTTCAAG 720
DB 661 CCAATGCTTCAAG 720
QY 721 AGAGATGCTTCAAG 780
DB 721 AGAGATGCTTCAAG 780
QY 781 AATAGCTTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 781 AATAGCTTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 840

QY 841 TATACATCTACTCTAACGACCGGAGAGAAAGTTCCCAAGACAGACTATCATTTG 900
 DB 841 TATACATCTACTCTAACGACCGGAGAGAAAGTTCCCAAGACAGACTATCATTTG 900
 QY 901 GAGGCTGACATTTGAACAGATCTTCAAAACATGAGACCTTACCAAGCTTTGGTGCAT 960
 DB 901 GAGGCTGACATTTGAACAGATCTTCAAAACATGAGACCTTACCAAGCTTTGGTGCAT 960
 QY 961 GCTGTTGGGCTCTCTCCCTGTTTAAAGCTTAAAGACACATCCAAAGCTTTCTAT 1020
 DB 961 GCTGTTGGGCTCTCTCCCTGTTTAAAGCTTAAAGACACATCCAAAGCTTTCTAT 1020
 QY 1021 ATGCATGACGAGACTATGAGACTACCTTTCTACGCTTTGACTGAGACTTCAACGCTTCC 1080
 DB 1021 ATGCATGACGAGACTATGAGACTACCTTTCTACGCTTTGACTGAGACTTCAACGCTTCC 1080
 QY 1081 GGTCTTATCAGAGAGAGTAAATGATGCTGCTTATCATGATCAAGACCTTTCCAGCTTC 1140
 DB 1081 GGTCTTATCAGAGAGAGTAAATGATGCTGCTTATCATGATCAAGACCTTTCCAGCTTC 1140
 QY 1141 CAGATTGATGTTATCTCGAAATACCTTACGAGGTTCTGACGCTTTGACTAGTCCGAA 1200
 DB 1141 CAGATTGATGTTATCTCGAAATACCTTACGAGGTTCTGACGCTTTGACTAGTCCGAA 1200
 QY 1201 ATGAAGATGCTCTTCTTCAAGTTGATATGTTGCTGAGAGATTCACAAAGTTGTTGG 1260
 DB 1201 ATGAAGATGCTCTTCTTCAAGTTGATATGTTGCTGAGAGATTCACAAAGTTGTTGG 1260
 QY 1261 GATGCTACTGCACTTCTCAGAGAGAGTACATATCAAACTGCAATGACAGACTTGG 1320
 DB 1261 GATGCTACTGCACTTCTCAGAGAGAGTACATATCAAACTGCAATGACAGACTTGG 1320
 QY 1321 CAGGAAGAGACAGAGATGAGTAACTTGAAGTGAAGACTTTTACGAGAGATG 1380
 DB 1321 CAGGAAGAGACAGAGATGAGTAACTTGAAGTGAAGACTTTTACGAGAGATG 1380
 QY 1381 TATGACCTTATGATGATGTTCCAGACCTTACACTCAGGTTGAGAGTGAAGGTTG 1440
 DB 1381 TATGACCTTATGATGATGTTCCAGACCTTACACTCAGGTTGAGAGTGAAGGTTG 1440
 QY 1441 TTTGAGGATGCTTACTTCAATCCTGATGTTGACTTGAACAATGGAAGACGTTAG 1500
 DB 1441 TTTGAGGATGCTTACTTCAATCCTGATGTTGACTTGAACAATGGAAGACGTTAG 1500
 QY 1501 TATGAGGCTTGAACCTTACTTTTGGGTAACCTGAACAGATTTGAAGGCAATGG 1560
 DB 1501 TATGAGGCTTGAACCTTACTTTTGGGTAACCTGAACAGATTTGAAGGCAATGG 1560
 QY 1561 TTTGAGGATCTTAAAGAGATCTTCAAAAACAAGCTTATCCCTTAAACCTTTAAG 1620
 DB 1561 TTTGAGGATCTTAAAGAGATCTTCAAAAACAAGCTTATCCCTTAAACCTTTAAG 1620
 QY 1621 GAGCCTAAGCAGCTAAATAGTAG 1644
 DB 1621 GAGCCTAAGCAGCTAAATAGTAG 1644
 RESULT 4
 AAT76552
 ID AAT76552 standard; DNA; 1801 BP.
 XX AAT76552;
 AC AAT76552;
 XX
 DT 12-SEP-1997 (first entry)
 XX
 XX DNA encoding hexose oxidase, an antimicrobial/antioxidant agent.
 XX
 XX Hexose oxidase; Chondrus crispus; marine algae; recombinant;
 KM antimicrobial; antioxidant; food preparation; dairy product; starch;
 KM beverage; animal feed; silage; sugar reduction; cosmetics; dental;
 KM toothpaste; dough; lactone production; ss.
 XX
 OS Chondrus crispus.

XX Key Location/Qualifiers
 FH CDS 84..1724
 FT /*tag= a
 FT
 PN W09640935-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 04-JUN-1996; 96MO-DK00238.
 XX
 PR 07-JUN-1995; 95US-0476910.
 XX
 PA (BIOT-) BIOTERNOLOGISK INST.
 XX
 PI Hansen OC, Strougaard P;
 XX
 DR WPI; 1997-052332/05.
 DR P-PSDB; AAM20076.
 XX
 PT Recombinant production of polypeptide having hexose oxidase activity
 PT - used in food preparations as antibacterial and antioxidant agent
 DR
 PS Claim 28; Page 103-104; 124pp; English.
 XX
 CC AAT76552 encodes hexose oxidase (HO) of the marine algae species
 CC Chondrus crispus. HO is useful in the production of food products, e.g.
 CC dairy products, starch-containing food products (dough) and non-dairy
 CC beverages. HO and active peptide fragments have antimicrobial and
 CC antioxidant properties and act by removing all the oxygen in a food
 CC packaging. HO and peptides of HO can also be used in an animal feed,
 CC especially silage. Further uses are to reduce, or analyse, the sugar
 CC content in a food, in the production of cosmetics, tooth care products
 CC or a pharmaceutical product and in lactone production. HO can be
 CC recombinantly produced in industrially appropriate quantities, and at
 CC a quality and purity level which renders the polypeptide suitable for
 CC industrial purposes.
 XX
 SQ Sequence 1801 BP; 441 A; 448 C; 492 G; 420 T; 0 other;
 Query Match 77.8%; Score 1279.2; DB 18; Length 1801;
 Best Local Similarity 86.1%; Pred. No. 0;
 Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;
 QY 1 ATGGCTACTCTTCCCAAAAGGACCCAGGTTCAATTTGATGAGTCAAGCTGGTACT 60
 DB 84 ATGGCTACTCTTCCCAAAAGGACCCAGGTTCAATTTGATGAGTCAAGCTGGTACT 143
 QY 61 CCAGACAAAGCTGACCCCAAGATTTGCATCCATGAAGCAAGTTTCAACAGAGATGATT 120
 DB 144 GCGGACAAAGCTGACCCCAAGATTTGCATCCATGAAGCAAGTTTCAACAGAGATGATT 203
 QY 121 GGTACCAACATGATTTGTTTACGTTTACGTTTACCTCCACAAAGTGTGTAAGTTG 180
 DB 204 GGAACATAATATGATTTGTTTATGTCGTACATCCCAAGTGTGTAAGTTG 263
 QY 181 GACAGAGCTATGAGAAAGTGTCTCCAGGTACCGTCAAGAAATGTTCTGAGTCACTG 240
 DB 264 GACCGGTATGAGAAAGTGTCTCCAGGTACCGTCAAGAAATGTTCTGAGTCACTG 323
 QY 241 TACGAAGACTTGTGTTTCCAGCAATGTTCAAGGCTATTTATCAAGTTTCTGTTGTT 300
 DB 324 TACGAAGACTTGTGTTTCCAGCAATGTTCAAGGCTATTTATCAAGTTTCTGTTGTT 383
 QY 301 GAATCTGTTACGAGCAGATAGAGTTACTTCTGTTCTTCCGTTGACCAACATGGGCT 360
 DB 384 GAGAGTGTATGAGCAGCAGATAGAGTTACTTCTGTTCTTCCGTTGACCAACATGGGCT 443
 QY 361 TCCTTCAAAAGCTTGTGTTTCAAGCAAGCAAGTATGTTTCCAGGTTGTTCTGTTACTCC 420
 DB 444 TCCTTCAAAAGCTTGTGTTTCAAGCAAGCAAGTATGTTTCCAGGTTGTTCTGTTACTCC 503
 QY 421 GTCCGTTTGGTGTGCACATTTGTCCGTGAGGTGACGATATTTTGGCCAGATTGACGCT 480

```

Db      504  GTGGGCTCGGTGGCCACATTTGTGGGCGAGGTGACCGCATTTTGGCCCTTGACATGGC 563
Qy      481  TTGGCAATCGATTTGGTATCCGGTGTGAGTTGTGCTTAAGCCAGCTTGACCGAAGAC 540
Db      564  CTCCCCCTGATTTGGCTCAGCGGCGTGAAGTGTCTTTAAAGCCAGTCTCACCGAAGAC 623
Qy      541  TCTGTTCTTAAGTACGTTTCAAGAGATTCCGAAGTAAACGACGGTGAAGTGTGTTGGCT 600
Db      624  TCGGTACTCAAGTATGTGCAAAAGATTCCGAAGGCAACGCGGAGGCTCTTTTGGCA 683
Qy      601  CACACTGTGAGGTGAGGTAACTTGTGTTATTCACCAATTCATCTTCAAGATTTG 660
Db      684  CACACAGGTGGCGGTGGCGAACTTGAATCATCAAAATCTACTCTTCAAGATTTG 743
Qy      661  CCATGTTCTCCAAAGAGTGTATGCTTTCACTTACCTTCTTGGGACGGTTGACT 720
Db      744  CCCATGTTCTCCAGGCGGTATGCAATCAATTTACCTTCAAGCTTGGACGGTTTACG 803
Qy      721  AGAGATCCTTGCAGAGATTTGTTGACTTAAGTACTTCAAGTTGGCTAGATGTGAG 780
Db      804  AGAGATCCTTGCAGAGATTTGTTGACTTAAGTACTTCAAGTTGGCTAGATGTGAG 863
Qy      781  AATATCTGTGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGGTTGTTATGACTTG 840
Db      864  AATACGTTGGCAAGTTCAAAATCTTCCATCAGGACGCGAAGAGTTGTATGACTTG 923
Qy      841  TATATCTCTTACTCTTAACGACCGCGAAGAGAAATTTCCCAAGACACACTATCATTTG 900
Db      924  TATATCTCTTACTCTTAACGACCGCGAAGAGAAATTTCCCAAGACCGCTACTATCATTTG 983
Qy      901  GAGCTGACATTTGAACAGATCTAACAAACATGCGAGCTTCAAGCTTGTGTGTCAT 960
Db      984  GAGCTGACATTTGAACAGATCTAACAAACATGCGAGCTTCAAGCTTGTGTGTCAT 1043
Qy      961  GCTGTTGGGCTCTCTTCTCTGTTAAGCTTAAGAAAGACACATCCAAAGACTTCTTAT 1020
Db      1044  GCTGTTGGGCTCTCTTCTCTGTTAAGCTTAAGAAAGACACATCCAAAGACTTCTTAT 1103
Qy      1021  ATGCATACAGAGATTAAGTAACTACCTTTTCAAGCTTGAAGTGAATCAAGCTTCC 1080
Db      1104  ATGCATACAGAGATTAAGTAACTACCTTTTCAAGCTTGAAGTGAATCAAGCTTCC 1163
Qy      1081  GGTCTTATCAGAGAGGTAAAGTAACTGCTTCAATGATCAAGACCTTCCAGACTTC 1140
Db      1164  GGTCTTATCAGAGAGGTAAAGTAACTGCTTCAATGATCAAGACCTTCCAGACTTC 1223
Qy      1141  CAGATTAATTTATCTGAAATACCTTACGTAGGTTCTCGACGGTTGACTGATGCCGAA 1200
Db      1224  CAGATTAATTTATCTGAAATACCTTACGTAGGTTCTCGACGGTTGACTGATGCCGAA 1283
Qy      1201  ATGAAGATGCTCTTCTTCAAGTGTATGTTGCTGTTGATGATTCACAAAGTGTGTTG 1260
Db      1284  ATGAAGATGCTCTTCTTCAAGTGTATGTTGCTGTTGATGATTCACAAAGTGTGTTG 1343
Qy      1261  GATCTACTGAGTGTGCTCAGAGAGATCATCATCAAACTGACGATCCAGACATTAAGT 1320
Db      1344  GATCTACTGAGTGTGCTCAGAGAGATCATCATCAAACTGACGATCCAGACATTAAGT 1403
Qy      1321  CAGGAAGAGCAAGATGACGTTAACTTGAAGTGAATGAGACTTTTTCGAGAGATG 1380
Db      1404  CAGGAAGAGCAAGATGACGTTAACTTGAAGTGAATGAGACTTTTTCGAGAGATG 1463
Qy      1381  TATAGACTTATGTTGTGTTCCAGACCTTAACACTAGGTTGAGAGTGAAGAGTGT 1440
Db      1464  TATAGACTTATGTTGTGTTCCAGACCTTAACACTAGGTTGAGAGTGAAGAGTGT 1523
Qy      1441  TTTGAGGATGCTACTTCACTACCTGATGTGATCTTGAACAACCTGGAAGAGCGTAA 1500
Db      1524  TTTGAGGATGCTACTTCACTACCTGATGTGATCTTGAACAACCTGGAAGAGCGTAA 1583
Qy      1501  TATGTTGCTTGAACCTTACTTTTGGTAACTGAAACAGATTGATCAAGCCAAATGG 1560

```

```

Db      1584  TATGTTGCCCTCGAAGCTTACTTTTGGTAACTGAAACCGGCTCATCAAGCCAAATG 1643
Qy      1561  TTGTGGATCTTAACGAGATCTTCAACAAACAAAGTCTATCTTAACTTAACTTAA 1620
Db      1644  TTGTGGATCTTAACGAGATCTTCAACAAACAAAGAGATCTTAACTTAACTTAA 1703
Qy      1621  GAGCTTAAGCAACTTAATATGATG 1644
Db      1704  GAGCCCAAGCAGCAAAATATGATG 1727

RESULT 5
AAZ94026/c
ID      AAZ94026 standard; DNA; 126 BP.
XX
AC      AAZ94026;
XX
DT      29-AUG-2000 (first entry)
XX
DE      Primer hox16 used in production of synthetic hexose oxidase gene.
XX
KM      Hexose oxidase; production; fermentation; synthetic; modification;
KM      prokaryote; eukaryote; ss.
XX
OS      Synthetic.
XX
PN      EP1008651-A2.
XX
PD      14-JUN-2000.
XX
PF      01-DEC-1999; 99EP-0204068.
XX
PR      09-DEC-1998; 98DK-0001630.
XX
PA      (BIOT-) BIOTEKNOLOGISK INST.
XX
XX      Strougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
XX      WPI; 2000-389309/34.
XX
PT      Nucleic acid fragment useful for producing large amounts of hexose
PT      oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT      modified by at least one codon
XX
PS      Example 2; Page 22; 42pp; English.
XX
CC      Wild type hexose oxidase can be produced by isolating a nucleotide
CC      sequence coding for hexose oxidase naturally produced by an organism
CC      and modifying the sequence so that the modified sequence is
CC      expressed, under identical conditions, at a level at least 10% higher
CC      than the non-modified sequence. The method is useful for producing
CC      hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC      of producing hexose oxidase were to isolate the enzyme from a source
CC      that naturally produces the enzyme, using host organisms which
CC      produce a relatively low level of expression. Therefore, industrial
CC      production of the enzyme is not feasible using this method. This is a
CC      much improved method which produces at least 250 mg enzyme per liter
CC      of fermentation medium. Twenty-eight primers (See AAZ94011-294038)
CC      with codon usage possibly optimised for expression in P. pastoris
CC      were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX
SQ      Sequence 126 BP; 33 A; 35 C; 23 G; 35 T; 0 other;

Query Match      7.7%; Score 126; DB 21; Length 126;
Best Local Similarity 100.0%; Pred No. 8.5e-28;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DB 66 GCAGTGTCTCAGAGAGATGATCATCAAACTGCAGTACCAGACATATCTGGCAGGAAGA 7
QY 1330 GACAAG 1335
|||||
DB 6 GACAAG 1

RESULT 6
ID AAS06168/c
AAS06168 standard; DNA; 126 BP.
XX AAS06168;
AC
XX
XX
DT 12-SEP-2001 (first entry)
XX
XX
DE PCR primer 6 used to clone synthetic HOX gene.
XX
XX HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW food manufacturing; beverage; detergent; baking; dough improving agent;
KW D-hexose:O2-oxidoreductase; EC 1.1.3.5; PCR primer; ss.
XX
XX Chondrus crispus.
OS Synthetic.
XX
XX WO200138544-A1.
PN
XX
XX 31-MAY-2001.
PD
XX
XX 24-NOV-2000; 2000MO-IB01886.
PF
XX
XX 24-NOV-1999; 99GB-0027801.
PR
XX
XX (DANI-) DANISCO AS.
PA
XX
PI Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
PI Zargahi MR;
XX
XX WPI; 2001-367695/38.
DR
XX
XX
PT Releasing soluble or membrane associated intracellular protein from a
PT cell for manufacturing food, comprises contacting the cell with a
PT membrane extracting composition and causing the protein to be released
PT in soluble form -
XX
XX
PS Claim 19; Fig 5; 108pp; English.

The sequence represents the PCR primer 6 used to clone synthetic hexose oxidase (D-hexose:O2-oxidoreductase, EC 1.1.3.5), (also referred to as HOX), gene. The native HOX gene was altered using site-directed mutagenesis in order to match the codon usage to known codon preferences of biotechnologically relevant yeasts, such as Pichia sp., to facilitate high level production in these organisms. The invention involves a method for releasing a soluble or membrane associated intracellular protein of interest (POI) from a cell involving contacting a cell comprising a soluble or membrane associated intracellular POI with a membrane extracting composition (1) and causing the POI to be released from the cell in a soluble form. The method is useful for releasing POI, such as an interleukin I receptor antagonist (IL-1ra) which involves contacting a transformed cell comprising IL-1ra with (1) and causing IL-1ra to be released from the transformed cell, in a soluble form. The method is also useful for screening mutated cells or transformed cells producing elevated levels of intracellular POI. The method is used to release a POI for manufacturing food products, such as beverages, preparation of detergents, and in baking as a dough improving agent. The method obtains a fast, specific and economically efficient extraction of a soluble or membrane associated intracellular POI without the use of conventional cell disruption techniques. The resulting cell extract contains less contaminating intracellular DNA and is relatively free of cell wall fragments. The intracellular POI can be recovered from a eukaryotic host organism such as yeast, before glycosylation takes place. The method can be used to prevent contact of intracellular POI with the extracellular growth medium.

SQ Sequence 126 BP; 33 A; 35 C; 23 G; 35 T; 0 other;
Query Match 7.7%; Score 126; DB 22; Length 126;
Best Local Similarity 100.0%; Pred. No. 8.5e-26;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1210 GCTCTTCTTCAGGTGATATGTTGGTGTGAGATTCAAGAAGTTGTTGGATGCTACT 1269
|||||
DB 126 GCTCTTCTTCAGGTGATATGTTGGTGTGAGATTCAAGAAGTTGTTGGATGCTACT 67
|||||

QY 1270 GCAGTGTCTCAGAGAGATGATCATCAAACTGCAGTACCAGACATATCTGGCAGGAAGA 1329
|||||
DB 66 GCAGTGTCTCAGAGAGATGATCATCAAACTGCAGTACCAGACATATCTGGCAGGAAGA 7
|||||

QY 1330 GACAAG 1335
|||||
DB 6 GACAAG 1

RESULT 7
ID AA294015
AA294015 standard; DNA; 120 BP.
XX
XX AA294015;
AC
XX
XX
DT 29-AUG-2000 (first entry)
XX
XX
DE Primer hoX5 used in production of synthetic hexose oxidase gene.
XX
XX Hexose oxidase; production; fermentation; synthetic; modification;
KW prokaryote; eukaryote; ss.
XX
XX Synthetic.
OS
XX
XX EPI008651-A2.
PN
XX
XX 14-JUN-2000.
PD
XX
XX
PF 01-DEC-1999; 99EP-0204068.
XX
XX
XX 09-DEC-1998; 98DK-0001630.
PR
XX
XX (BIOT-) BIOTEKNOLOGISK INST.
PA
XX
PI Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
PI
XX
XX WPI; 2000-389309/34.
DR
XX
XX

Nucleic acid fragment useful for producing large amounts of hexose oxidase comprises nucleotide sequence encoding hexose oxidase that is modified by at least one codon

Example 2; Page 20; 42pp; English.

Wild type hexose oxidase can be produced by isolating a nucleotide sequence coding for hexose oxidase naturally produced by an organism and modifying the sequence so that the modified sequence is an organism expressed, under identical conditions, at a level at least 10% higher than the non-modified sequence. The method is useful for producing hexose oxidase in prokaryotic and eukaryotic cells. Previous methods of producing hexose oxidase were to isolate the enzyme from a source that naturally produces the enzyme, using host organisms which produce a relatively low level of expression. Therefore, industrial production of the enzyme is not feasible using this method. This is a much improved method which produces at least 250 mg enzyme per liter of fermentation medium. Twenty-eight primers (See AA294011-294038) with codon usage possibly optimised for expression in P. pastoris were used to produce the synthetic hexose oxidase gene (See AA294039).

SQ Sequence 120 BP; 26 A; 26 C; 31 G; 37 T; 0 other;
Query Match 7.3%; Score 120; DB 21; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.4e-26;

Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 ATCAAGCTTACTGTTGTTGATGCTGATGAGAGATGAGTACTTGGCTCT 339
 Db 1 ATCAAGCTTACTGTTGTTGATGCTGATGAGAGATGAGTACTTGGCTCT 60

QY 340 TCCGGTACACCACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCGTAGAGTTTG 399
 Db 61 TCCGGTACACCACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCGTAGAGTTTG 120

RESULT 10
 AAS06166/c
 ID AAS06166 standard; DNA; 120 BP.
 XX AAS06166;
 AC AAS06166;
 XX 12-SEP-2001 (first entry)
 DT PCR primer 4 used to clone synthetic HOX gene.
 DE
 XX
 XX HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
 KM food manufacturing; beverage; detergent; baking; dough improving agent;
 KW D-hexose:O2-oxidoreductase; EC 1.1.3.5; PCR primer; ss.
 XX
 OS Chondrus crispus.
 OS Synthetic.
 OS
 PN WO200138544-A1.
 XX
 PD 31-MAY-2001.
 PF 24-NOV-2000; 2000WO-IB01886.
 XX
 PR 24-NOV-1999; 99GB-0027801.
 XX
 PA (DANI-) DANISCO AS.
 XX
 PI Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
 XX Zargahi MR;
 XX WPI; 2001-367695/38.
 DR
 XX
 XX
 PT Releasing soluble or membrane associated intracellular protein from a
 PT cell for manufacturing food, comprises contacting the cell with a
 PT membrane extracting composition and causing the protein to be released
 PT in soluble form -
 PT
 PS Claim 19; Fig 5; 108pp; English.
 XX
 XX The sequence represents the PCR primer 4 used to clone synthetic
 CC hexose oxidase (D-hexose:O2-oxidoreductase, EC 1.1.3.5), (also referred to
 CC as HOX), gene. The native HOX gene was altered using site-directed
 CC mutagenesis in order to match the codon usage to known codon preferences
 CC of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
 CC high level production in these organisms. The invention involves a method
 CC for releasing a soluble or membrane associated intracellular protein of
 CC interest (POI) from a cell involving contacting a cell comprising a
 CC soluble or membrane associated intracellular POI with a membrane
 CC extracting composition (1) and causing the POI to be released from the
 CC cell in a soluble form. The method is useful for releasing POI, such as
 CC an interleukin I receptor antagonist (IL-1ra) which involves contacting a
 CC transformed cell comprising IL-1ra with (1) and causing IL-1ra to be
 CC released from the transformed cell, in a soluble form. The method is also
 CC useful for screening mutated cells or transformed cells producing
 CC elevated levels of intracellular POI. The method is used to release a POI
 CC for manufacturing food products, such as beverages, preparation of
 CC detergents, and in baking as a dough improving agent. The method obtains
 CC a fast, specific and economically efficient extraction of a soluble or
 CC membrane associated intracellular POI without the use of conventional
 CC cell disruption techniques. The resulting cell extract contains less
 CC contaminating intracellular DNA and is relatively free of cell wall
 CC fragments. The intracellular POI can be recovered from a eukaryotic host

CC organism such as yeast, before glycosylation takes place. The method can
 CC be used to prevent contact of intracellular POI with the extracellular
 CC growth medium.
 CC
 XX Sequence 120 BP; 33 A; 25 C; 29 G; 33 T; 0 other;

QY 1027 GACGAGACTAGTACCTCCCTTTCTTACCGTTTGAAGACTATCAACGGTCCGCTCT 1086
 Db 120 GACGAGACTAGTACCTCCCTTTCTTACCGTTTGAAGACTATCAACGGTCCGCTCT 61

QY 1087 AATCAGAGAGTATGATGATGCTGCTTATCATGATCAAGACTTCCAGACTTCCAGATT 1146
 Db 60 AATCAGAGAGTATGATGATGCTGCTTATCATGATCAAGACTTCCAGACTTCCAGATT 1

RESULT 11
 AA294019
 ID AA294019 standard; DNA; 118 BP.
 XX AA294019;
 AC AA294019;
 XX 29-AUG-2000 (first entry)
 DT
 XX
 DE Primer hox9 used in production of synthetic hexose oxidase gene.
 XX
 XX Hexose oxidase; production; fermentation; synthetic; modification;
 KM prokaryote; eukaryote; ss.
 XX
 OS Synthetic.
 OS
 PN EP1008651-A2.
 XX
 PD 14-JUN-2000.
 PF 01-DEC-1999; 99EP-0204068.
 XX
 PR 09-DEC-1998; 98DK-0001630.
 XX
 PA (BIOT-) BIOTEKNOLOGISK INST.
 XX
 PI Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
 XX WPI; 2000-389309/34.
 DR
 XX
 XX Nucleic acid fragment useful for producing large amounts of hexose
 PT oxidase comprises nucleotide sequence encoding hexose oxidase that is
 PT modified by at least one codon
 PT
 PS Example 2; Page 21; 42pp; English.
 XX
 XX Wild type hexose oxidase can be produced by isolating a nucleotide
 CC sequence coding for hexose oxidase naturally produced by an organism
 CC and modifying the sequence so that the modified sequence is
 CC expressed, under identical conditions, at a level at least 10% higher
 CC than the non-modified sequence. The method is useful for producing
 CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
 CC of producing hexose oxidase were to isolate the enzyme from a source
 CC that naturally produces the enzyme, using host organisms which
 CC produce a relatively low level of expression. Therefore, industrial
 CC production of the enzyme is not feasible using this method. This is a
 CC much improved method which produces at least 250 mg enzyme per liter
 CC of fermentation medium. Twenty-eight primers (See AA294011-294038)
 CC with codon usage possibly optimised for expression in P. pastoris
 CC were used to produce the synthetic hexose oxidase gene (See AA294039).
 XX
 XX Sequence 118 BP; 30 A; 27 C; 22 G; 39 T; 0 other;

Query Match 7.2%; Score 118; DB 21; Length 118;
 Best Local Similarity 100.0%; Pred. No. 2.1e-25;

	Matches	118;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	
Oy		639	CAATACACTCTTCAGGATTGGCCAAATGTCTTCCAGAGGTGCATCGCTTTCACTTACA	698						
Dd		1	CAAAATACACTCTTCAGGATTGGCCAAATGTCTTCCAGAGGTGCATCGCTTTCACTTACA	60						
Oy		699	CTTCTCTTGGAACGGTTTCACATAGAGATGCCCTTGCAGAATTTGTTGACTAATGACTTC	756						
Dd		61	CTTCTCTTGGAACGGTTTCACATAGAGATGCCCTTGCAGAATTTGTTGACTAATGACTTC	118						
	RESULT 12									
ID	AAS06161									
XX	AAS06161 standard; DNA; 118 BP.									
AC										
XX	AAS06161;									
XX										
DT	12-SEP-2001 (first entry)									
XX										
DE	PCR primer hox9a used to clone synthetic HOX gene.									
XX										
KM	HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;									
KM	food manufacturing; beverage; detergent; baking; dough improving agent;									
KM	D-hexose:O2-oxidoreductase; EC 1.1.3.5; PCR primer; hox9a; ss.									
OS	Chondrus crispus.									
OS	Synthetic.									
XX										
PN	WO200138544-A1.									
XX										
FD	31-MAY-2001.									
XX										
PF	24-NOV-2000; 2000WO-IB01886.									
XX										
PR	24-NOV-1999; 99GB-0027801.									
XX										
PA	(DANI-) DANISCO AS.									
Pt	Johansen CL, Kjaerulf S, Madrid SM, Pedersen H, Poulsen CH,									
Pt	Zargani MR;									
DR	WPI; 2001-367695/38.									
XX										
PT	Releasing soluble or membrane associated intracellular protein from a									
PT	cell for manufacturing food, comprises contacting the cell with a									
PT	membrane extracting composition and causing the protein to be released									
XX	in soluble form -									
PS	Claim 19; Fig 5; 108pp; English.									
XX										
CC	The sequence represents the PCR primer hox9a used to clone synthetic									
CC	hexose oxidase (D-hexose:O2-oxidoreductase, EC 1.1.3.5), (also referred to									
CC	as HOX), gene. The native HOX gene was altered using site-directed									
CC	mutagenesis in order to match the codon usage to known codon preferences									
CC	of biotechnologically relevant yeasts, such as Pichia sp., to facilitate									
CC	high level production in these organisms. The invention involves a method									
CC	for releasing a soluble or membrane associated intracellular protein of									
CC	interest (POI) from a cell involving contacting a cell comprising a									
CC	soluble or membrane associated intracellular POI with a membrane									
CC	extracting composition (I) and causing the POI to be released from the									
CC	cell in a soluble form. The method is useful for releasing POI, such as									
CC	an interleukin I receptor antagonist (IL-1ra) which involves contacting a									
CC	transformed cell comprising IL-1ra with (I) and causing IL-1ra to be									
CC	released from the transformed cell, in a soluble form. The method is also									
CC	useful for screening mutated cells or transformed cells producing									
CC	elevated levels of intracellular POI. The method is used to release a POI									
CC	for manufacturing food products, such as beverages, preparation of									
CC	detergents, and in baking as a dough improving agent. The method obtains									
CC	a fast, specific and economically efficient extraction of a soluble or									
CC	membrane associated intracellular POI without the use of conventional									
CC	cell disruption techniques. The resulting cell extract contains less									
CC	contaminating intracellular DNA and is relatively free of cell wall									
CC	fragments. The intracellular POI can be recovered from a									

CC	organism such as yeast, before glycosylation takes place. The method can
CC	be used to prevent contact of intracellular POI with the extracellular
CC	growth medium.
XX	
SQ	Sequence 118 BP; 30 A; 27 C; 22 G; 39 T; 0 other;
OY	
Db	Query Match Best Local Similarity 100.0%; Pred.No.2,1e-25; Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps
Oy	
Db	639 CAANAATCTACTTTCAAGAGATTGGCATTGTCCTCCAAAGAGGTGCATCGCTTCTTAACCTACA 639 1 CAATAATCTACTTTCAAGAGATTGGCATTGTCCTCCAAAGAGGTGCATCGCTTCTTAACCTACA 600
Db	699 CTTCCTCTTGSGAGCGGTTTCACCTAGAGATGCCCTTGCAAGATTTGTTGACTAATGACTTTC 756 61 CTTCCTCTTGSGAGCGGTTTCACCTAGAGATGCCCTTGCAAGATTTGTTGACTAATGACTTTC 118
RESULT 13	
ID	AAZ94029 standard; DNA; 117 BP.
XX	
AC	AAZ94029;
DT	29-AUG-2000 (first entry)
DE	Primer hox19 used in production of synthetic hexose oxidase gene.
XX	
KM	Hexose oxidase; production; fermentation; synthetic; modification; prokaryote; eukaryote; ss.
OS	Synthetic.
PN	EPI008651-A2.
PD	14-JUN-2000.
PF	01-DEC-1999; 99EP-0204068.
PR	09-DEC-1998; 98DK-0001630.
PA	(BIOT-) BIOTEKNOLOGISK INST.
XX	
Pt	Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
XX	
DR	WPI; 2000-389309/34.
XX	
PT	Nucleic acid fragment useful for producing large amounts of hexose
PT	oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT	modified by at least one codon
XX	
PS	Example 2; Page 23; 42pp; English.
XX	
CC	Wild type hexose oxidase can be produced by isolating a nucleotide
CC	sequence coding for hexose oxidase naturally produced by an organism
CC	and modifying the sequence so that the modified sequence is
CC	expressed under identical conditions, at a level at least 10% higher
CC	than the non-modified sequence. The method is useful for producing
CC	hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC	of producing hexose oxidase were to isolate the enzyme from a source
CC	that naturally produces the enzyme, using host organisms which
CC	produce a relatively low level of expression. Therefore, industrial
CC	production of the enzyme is not feasible using this method. This is a
CC	much improved method which produces at least 250 mg enzyme per liter
CC	of fermentation medium. Twenty-eight primers (See AAZ94011-Z94038)
CC	with codon usage possibly optimised for expression in P. pastoris
CC	were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX	
SQ	Sequence 117 BP; 36 A; 23 C; 26 G; 32 T; 0 other;
Query Match	7.1%; Score 117; DB 21; Length 117;
Best Local Similarity	100.0%; Pred.No.4,3e-25;

CC elevated levels of intracellular POI. The method is used to release a POI
CC for manufacturing food products, such as beverages, preparation of
CC detergents, and in baking as a dough improving agent. The method obtains
CC a fast, specific and economically efficient extraction of a soluble or
CC membrane associated intracellular POI without the use of conventional
CC cell disruption techniques. The resulting cell extract contains less
CC contaminating intracellular DNA and is relatively free of cell wall
CC fragments. The intracellular POI can be recovered from a eukaryotic host
CC organism such as yeast, before glycosylation takes place. The method can
CC be used to prevent contact of intracellular POI with the extracellular
CC growth medium.

	Query Match	Similarity	Score	DB	Length
Best Local Match	Similarity	99.28	Pred.	6.5e-25	
Matches	117	Conservative	0	Mismatches	1
				Indels	0
				Gaps	0

	Query	Match	Similarity	Score	DB	Length
Qy	732	GCAAGATTGTTGACTAAGTCTTCAAGTGGCTGATGTGATTTGGAAGAACTACTGTTGG	791			
Db	118	GCAAGATTGTTGACTAAGTCTTCAAGTGGCTGATGTGATTTGGAAGAACTACTGTTGG	59			
Qy	792	TAAGTTCGAAATCTTCCACCAAGCAGCTGAGAGATTGTTATGATCTTGTATACATCC	849			
Db	58	TAAGTTCGAAATCTTCCACCAAGCAGCTGAGAGATTGTTATGATCTTGTATACATCC	1			

Search completed: June 15, 2003, 02:56:53
Job time : 389 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 02:45:32 ; Search time 4309 Seconds
(without alignments)

11103.507 Million cell updates/sec

Title: US-09-998-284-1

Perfect score: 1644

Sequence: 1 atgagctacttgcacacaaa.....ctaaagcagactaaatagtag 1644

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GeneBml:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rod:*
36: em_hcg_man:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1644	100.0	1644	6 AX155104	AX155104 Sequence
2	1638	99.6	1638	12 SC0294935	AJ294935 Synthetic
3	1280.8	77.9	1881	8 CCUB98770	U89770 Chondrus cr
4	1279.2	77.8	1801	6 AR159722	AR159722 Sequence
5	1279.2	77.8	1801	6 AX030569	AX030569 Sequence
6	126	7.7	126	6 AX155088	AX155088 Sequence
7	120	7.3	120	6 AX155088	AX155088 Sequence
8	120	7.3	120	6 AX155097	AX155097 Sequence
9	118	7.2	118	6 AX155092	AX155092 Sequence
10	117	7.1	117	6 AX155102	AX155102 Sequence
11	116.4	7.1	116	6 AX155093	AX155093 Sequence
12	116	7.1	116	6 AX155093	AX155093 Sequence
13	111	6.8	111	6 AX155101	AX155101 Sequence
14	109	6.6	109	6 AX155089	AX155089 Sequence
15	109	6.6	109	6 AX155089	AX155089 Sequence
16	108	6.6	108	6 AX155098	AX155098 Sequence
17	108	6.6	108	6 AX155100	AX155100 Sequence
18	107	6.5	107	6 AX155085	AX155085 Sequence
19	107	6.5	107	6 AX155085	AX155085 Sequence
20	106	6.4	106	6 AX155086	AX155086 Sequence
21	106	6.4	106	6 AX155087	AX155087 Sequence
22	102	6.2	102	6 AX155095	AX155095 Sequence
23	92.8	5.6	96	6 AX155094	AX155094 Sequence
24	85.4	5.2	10668	1 AE013658	AE013658 Yersinia
25	85.4	5.2	214050	1 AJ414142	AJ414142 Yersinia
26	81.8	5.0	90	6 AX155096	AX155096 Sequence
27	79.6	4.8	24225	1 SC2C4	AL512902 Streptomy
28	77.2	4.7	215050	1 AL646057	AL646057 Ralstonia
29	69.2	4.2	25883	1 AB008466	AB008466 Streptomy
30	63	3.8	26195	1 SC1G7	AL591083 Streptomy
31	60	3.6	78	6 AX155103	AX155103 Sequence
32	60	3.6	1620	6 AX089430	AX089430 Sequence
33	60	3.6	45624	6 AX089419	AX089419 Sequence
34	60	3.6	50000	6 AX089416	AX089416 Sequence
35	60	3.6	80161	1 AY007564	AY007564 Saccharop
36	56.8	3.5	80161	6 AR165018	AR165018 Sequence
37	56	3.4	2088	1 SPUB0222	U80222 Streptomyce
38	43.6	2.7	61	6 AX155084	AX155084 Sequence
39	42.6	2.6	1141	6 AX083744	AX083744 Sequence
40	42.6	2.6	1738	8 AY051000	AY051000 Arabidops
41	42	2.6	89473	8 T1K7	AC013427 Sequence
42	42	2.6	1639	8 AY113892	AY113892 Arabidops
43	42	2.6	1785	8 AF360332	AF360332 Arabidops
44	42	2.6	51939	9 AC099341	AC099341 Homo sapi
45	41.8	2.5	91566	8 F25C20	AC007296 Arabidops
			1584	6 AX412693	AX412693 Sequence

ALIGNMENTS

RESULT 1
LOCUS AX155104 1644 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 22 from Patent WO0138544.
ACCESSION AX155104
VERSION AX155104.1 GI:14536652
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
AUTHORS 1 (bases 1 to 1644)
Johansen,C.L., Kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
Foulsen,C.H. and Zargahl,M.R.
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 22 31-MAY-2001;

Pred. No. is the number of results predicted by chance to have a

FEATURES DANISCO A/S (DK)
Location/Qualifiers
source 1..1644
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Nucleotide sequence"
BASE COUNT 430 a 337 c 404 g 473 t
ORIGIN
Query Match 100.0%; Score 1644; DB 6; Length 1644;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTACTTGGCCAAAGAGACCAGTTACATTTGTAAGTCAACGCTGTACT 60
DB 1 ATGGCTACTTGGCCAAAGAGACCAGTTACATTTGTAAGTCAACGCTGTACT 60
QY 61 CCAGACCAAGCCTGACCAAGATTGCCATCATGAGCAAGTTCAACAGAAATGATT 120
DB 61 CCAGACCAAGCCTGACCAAGATTGCCATCATGAGCAAGTTCAACAGAAATGATT 120
QY 121 GGTAACCAATCGATTGTTAGTGTGTTACATCCACAAGGTGTTGTAAGTCTTG 180
DB 121 GGTAACCAATCGATTGTTAGTGTGTTACATCCACAAGGTGTTGTAAGTCTTG 180
QY 181 GACAGAGCTATGAAAAGTCTCCAGTACCGTCAAGATCGTTCTGGTGTCACTGT 240
DB 181 GACAGAGCTATGAAAAGTCTCCAGTACCGTCAAGATCGTTCTGGTGTCACTGT 240
QY 241 TACGAAGCTTCTGTTTGAAGATGTGTCAGAGCTATTAACAAGTTACTGTTGGTT 300
DB 241 TACGAAGCTTCTGTTTGAAGATGTGTCAGAGCTATTAACAAGTTACTGTTGGTT 300
QY 301 GAATCTGGTTACGACGAGATAGAGTTACTGTTCTCCGGTGAACCACTGGGGT 360
DB 301 GAATCTGGTTACGACGAGATAGAGTTACTGTTCTCCGGTGAACCACTGGGGT 360
QY 361 TCCCTTCAAGACCTTGTTCAGAGCAACGATAGATTTTGCAGGTGTTCTGTACTCC 420
DB 361 TCCCTTCAAGACCTTGTTCAGAGCAACGATAGATTTTGCAGGTGTTCTGTACTCC 420
QY 421 GTGCGTTGGGTGTCAATTTGCGTGAAGTGAAGTATTTGGCCAAATGTCAGGT 480
DB 421 GTGCGTTGGGTGTCAATTTGCGTGAAGTGAAGTATTTGGCCAAATGTCAGGT 480
QY 481 TTGCGAGTCAATTTGTTATCGGTTGTAAGTGTGTTAAGCCAGCTTGACCGAAGAC 540
DB 481 TTGCGAGTCAATTTGTTATCGGTTGTAAGTGTGTTAAGCCAGCTTGACCGAAGAC 540
QY 541 TCTGTTCTTAAGTCACTTCAAGAGATTCGAAGTGAACGAGTGTGTTGGGCT 600
DB 541 TCTGTTCTTAAGTCACTTCAAGAGATTCGAAGTGAACGAGTGTGTTGGGCT 600
QY 601 CACACTGTGAGAGTGAAGTGAATCTGATTAATCAACAAATCTCTCAAGATTTG 660
DB 601 CACACTGTGAGAGTGAAGTGAATCTGATTAATCAACAAATCTCTCAAGATTTG 660
QY 661 CCATGCTCTCAAGAGTGTCACTTCACTTAAGTCTCTGGAAGCGTTTACT 720
DB 661 CCATGCTCTCAAGAGTGTCACTTCACTTAAGTCTCTGGAAGCGTTTACT 720
QY 721 AGAGATGCTTGAAGATTTGTTGATTAAGTCAAGTGTGCTAAGTGTGAAG 780
DB 721 AGAGATGCTTGAAGATTTGTTGATTAAGTCAAGTGTGCTAAGTGTGAAG 780
QY 781 AATAGTGTGTAAGTCAATCTTCCACCAAGCACTGAAGAGTTGTAATGACTTG 840
DB 781 AATAGTGTGTAAGTCAATCTTCCACCAAGCACTGAAGAGTTGTAATGACTTG 840
QY 841 TATACATCTTACTTAACGAGCGCGAAGAGAGATGTCACCAAGCACTATCAATTTG 900
DB 841 TATACATCTTACTTAACGAGCGCGAAGAGAGATGTCACCAAGCACTATCAATTTG 900

QY 901 GAGCTGACATGTAACAGATCTACAAACATGCGAGCTTACCAAGCTTGTGTCAT 960
DB 901 GAGCTGACATGTAACAGATCTACAAACATGCGAGCTTACCAAGCTTGTGTCAT 960
QY 961 GCTGTTGGGCTCCTTTCCCTGTTAGCCTAGAAAGACACATCCAAAGCTTTAT 1020
DB 961 GCTGTTGGGCTCCTTTCCCTGTTAGCCTAGAAAGACACATCCAAAGCTTTAT 1020
QY 1021 ATGCATGACGAGACTATGACTACCTTTCTACAGCTTTGATGAGACTATCAACGTTCC 1080
DB 1021 ATGCATGACGAGACTATGACTACCTTTCTACAGCTTTGATGAGACTATCAACGTTCC 1080
QY 1081 GGTCCTATTCAGAGCTAAGTAACTGCTTCAATCATGATCAAGACTTCCAGACTTC 1140
DB 1081 GGTCCTATTCAGAGCTAAGTAACTGCTTCAATCATGATCAAGACTTCCAGACTTC 1140
QY 1141 CAGATTGATTTATCTGAAATACCTTACTGAGGTTCTGAGGTTGATGCGGAA 1200
DB 1141 CAGATTGATTTATCTGAAATACCTTACTGAGGTTCTGAGGTTGATGCGGAA 1200
QY 1201 ATGAAGATGCTCTTCTTCAAGTTGATATGTTGCGTGTGAGATTCAAGAGTTGTTGG 1260
DB 1201 ATGAAGATGCTCTTCTTCAAGTTGATATGTTGCGTGTGAGATTCAAGAGTTGTTGG 1260
QY 1261 GATGCTACTGCAAGTGTCTCAGAGAGTACATCATCAAACTGCAAGTCAAGACTGCG 1320
DB 1261 GATGCTACTGCAAGTGTCTCAGAGAGTACATCATCAAACTGCAAGTCAAGACTGCG 1320
QY 1321 CAGGAAGAAACAAGATGCACTTAATGATGATTTAGAGACTTTTACGAGAGATG 1380
DB 1321 CAGGAAGAAACAAGATGCACTTAATGATGATTTAGAGACTTTTACGAGAGATG 1380
QY 1381 TATGACCTTATGATGATGTTCCAGACCTTAACACTCAGGTTGAGAGTGAAGTGT 1440
DB 1381 TATGACCTTATGATGATGTTCCAGACCTTAACACTCAGGTTGAGAGTGAAGTGT 1440
QY 1441 TTGAGGATGCTACTTCAACTACCTGATGTTGATGATCAACTGGAAGACGTTAG 1500
DB 1441 TTGAGGATGCTACTTCAACTACCTGATGTTGATGATCAACTGGAAGACGTTAG 1500
QY 1501 TATGTCCTTGAACCTTATCTTTGGTGAACCTGAACAGATGATGATCAAGCCAAATG 1560
DB 1501 TATGTCCTTGAACCTTATCTTTGGTGAACCTGAACAGATGATGATCAAGCCAAATG 1560
QY 1561 TTGTGGATCTTAACAGATCTTCAACAACAAACATCTATCCCTTAACCTTTAG 1620
DB 1561 TTGTGGATCTTAACAGATCTTCAACAACAAACATCTATCCCTTAACCTTTAG 1620
QY 1621 GAGCCTAAGCAGCTAATATGATG 1644
DB 1621 GAGCCTAAGCAGCTAATATGATG 1644
RESULT 2
SCO294935 1638 bp DNA linear SYN 09-OCT-2000
LOCUS Synthetic construct partial shox gene for hexose oxidase.
DEFINITION
ACCESSION AJ294935
VERSION AJ294935.1 GI:10798685
KEYWORDS hexose oxidase; SHOX gene.
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1638)
AUTHORS Wolff, A., Hansen, O.C., Poulsen, U., Madrid, S. and Stougaard, P.
TITLE Recombinant production of hexose oxidase from the red alga Chondrus crispus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1638)
AUTHORS Stougaard, P.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2000) Stougaard P., Enzyme Technology, Biotechnological Institute, Kogle Alle 2, DK-2970 Hoersholm,

FEATURES		DENMARK	
source	Location/Qualifiers		
gene	1..1638	/organism="synthetic construct"	
	/db_xref="taxon:32630"	/note="Chondrus crispus sequence with codon usage optimized to Pichia pastoris"	
	1..1638	/gene="SHOX"	
	1..1638	/gene="SHOX"	
	/EC_number="1.1.3.5"	/note="86% sequence homology to native gene; 100% identity with native aa sequence"	
CDS	/codon_start=1	/evidence=experimental	
	/transl_table=11	/product="hexose oxidase"	
	/protein_id="CAC12830.1"	/db_xref="GI:10798866"	
	/translation="MATLPQKDPGYIVDVNAGTPDKPRLPLSMKOGFNRRWIGTNI DPAVVYVTPQAGCTALDRAMEKSPGVTVLGGHCYEDPVEDCEKAIINVTGLVES GYDDRGVYVSSGDTWGSFKTLFRDHGVVLPBGSCTSVLGGHIVGGSDILARIHG LPLDMLSGEVVYVVKPVLTEDSVLKVYHKDSEGDGLFWAHGSGGNGFGLITKTYFK DLPMSPRGIASNLHPSWDGFTKDALQDLITKFKLARCDMKNTVKGFIQQAEEF VWLYTYSNDADREVAODRHYHLEADIEQIYKTCPTALGSHAGMADFPVPRKH TSKTSYMDETMDYFVALTEINSGSPNORCKSAVMIKDPDPDIDVIMKYLEEV PDGLTSAEMKDALIDVDMFGELHKVVMATAVAOREVYIKILOYOTVMOEEDKAVNL KNIRDPYEMERYPGGVPPNPVQESKGFVFCGCTFNYRVDVLDLNNKNGKYGALBIYF LGNLNILIAKMLMDPNEILFTNKSQIPIYPLKEPKQTK"		
	BASE COUNT	428 a	337 c 402 g 471 t
	ORIGIN		
	Query Match	99.6%; Score 1638; DB 12; Length 1638;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 1638; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
	QY	1	ATGGCTACTTTGGCCAAAGAGACCCAGGTTACATTTGTTAGCGTCAACGCTGGTACT
DB	1	ATGGCTACTTTGGCCAAAGAGACCCAGGTTACATTTGTTAGCGTCAACGCTGGTACT	60
QY	61	CCAGACAAAGCTGACCCCAAGATTGCCATCAGTAAGCAAGGTTTCAACAGAAAGTGAAT	120
DB	61	CCAGACAAAGCTGACCCCAAGATTGCCATCAGTAAGCAAGGTTTCAACAGAAAGTGAAT	120
QY	121	GGTACCAATCGATTTCGTTTACGTCGTTTACATCTCCACAAAGTGTGTACTGCTTTG	180
DB	121	GGTACCAATCGATTTCGTTTACGTCGTTTACATCTCCACAAAGTGTGTACTGCTTTG	180
QY	181	GACAGAGCTATGAGAAAGTGTTCTCCAGGTACCGTCAAGATCGTTTCGGTGTGCACTGT	240
DB	181	GACAGAGCTATGAGAAAGTGTTCTCCAGGTACCGTCAAGATCGTTTCGGTGTGCACTGT	240
QY	241	TACGAAGACTTCGTTTTCGACGAATGTGTCAGAGCTATTATCAACGTTACTGCTTTG	300
DB	241	TACGAAGACTTCGTTTTCGACGAATGTGTCAGAGCTATTATCAACGTTACTGCTTTG	300
QY	301	GAATCTGGTTACGACGACGATAGAGTTACTTCTCTTCGCGTGACACCACTGGGGT	360
DB	301	GAATCTGGTTACGACGACGATAGAGTTACTTCTCTTCGCGTGACACCACTGGGGT	360
QY	361	TCCTTCAAGACCTTGTTCAGAGACCAACGAGTAGAGTTTTCGAGGTGTTCTGTACTCC	420
DB	361	TCCTTCAAGACCTTGTTCAGAGACCAACGAGTAGAGTTTTCGAGGTGTTCTGTACTCC	420
QY	421	GTCGTTTGGGTGTCACATTTGTCGATGAGGTGACCGTAATTTTGGCCAGATTGACGGT	480
DB	421	GTCGTTTGGGTGTCACATTTGTCGATGAGGTGACCGTAATTTTGGCCAGATTGACGGT	480
QY	481	TTGGCCAGTGCATTGTATCCGGTGTGAAGTTGTGTTAAGCAGCTTGACCGAAGAC	540
DB	481	TTGGCCAGTGCATTGTATCCGGTGTGAAGTTGTGTTAAGCAGCTTGACCGAAGAC	540

QY	541	TTGTGTTCTTAAGTACGTTTCAACAAGATTCGGAAGTAAACAACGGTAGTGTCTTTGGGCT	600
DB	541	TCTGTTCTTAAGTACGTTTCAACAAGATTCGGAAGTAAACAACGGTAGTGTCTTTGGGCT	600
QY	601	CACACTGGATGAGGAGGAGGTAACTTGCGTATTTATCAACAAATCTACTTCAAGAAATTG	660
DB	601	CACACTGGATGAGGAGGAGGTAACTTGCGTATTTATCAACAAATCTACTTCAAGAAATTG	660
QY	661	CCAAATGTCACAAAGAGTGTGATCGCTTCTTAATTACACTTCTCTTGGGACGTTTCACT	720
DB	661	CCAAATGTCACAAAGAGTGTGATCGCTTCTTAATTACACTTCTCTTGGGACGTTTCACT	720
QY	721	AGAGATGCCCTTGCAAGATTGTTGTAAGTAACTTCAAGTTGGCTAGATGATTTGGAAG	780
DB	721	AGAGATGCCCTTGCAAGATTGTTGTAAGTAACTTCAAGTTGGCTAGATGATTTGGAAG	780
QY	781	AATACGTTGGTAATTCCTCAAACTTTCACCAAGAGGTGAAGTTGTTATGACTTG	840
DB	781	AATACGTTGGTAATTCCTCAAACTTTCACCAAGAGGTGAAGTTGTTATGACTTG	840
QY	841	TATATATCTCTACTCTTAACGACGCGCAGAGAGAAGTTGGCCAAACAGACATATCATTTG	900
DB	841	TATATATCTCTACTCTTAACGACGCGCAGAGAGAAGTTGGCCAAACAGACATATCATTTG	900
QY	901	GAGGCTGACATTGAACAGATCTCAAAAATGCGAGCTTACCAAACTCTTGGTGTGAT	960
DB	901	GAGGCTGACATTGAACAGATCTCAAAAATGCGAGCTTACCAAACTCTTGGTGTGAT	960
QY	961	GCTGTTGGGCTCCCTTCCCTGTTAGACTTGAAGAAAGACACATCCAAAGCTTCTTAT	1020
DB	961	GCTGTTGGGCTCCCTTCCCTGTTAGACTTGAAGAAAGACACATCCAAAGCTTCTTAT	1020
QY	1021	ATGCATGACGAGACTATGAACTACCCCTTTCACGCTTGAAGTGAAGTCAACGATTC	1080
DB	1021	ATGCATGACGAGACTATGAACTACCCCTTTCACGCTTGAAGTGAAGTCAACGATTC	1080
QY	1081	GGTCTCTAATCAGAGAGGTAAAGTCAAGTCTGTTACATGATCAAGACTTTCACAGCTTC	1140
DB	1081	GGTCTCTAATCAGAGAGGTAAAGTCAAGTCTGTTACATGATCAAGACTTTCACAGACTTC	1140
QY	1141	CAGATTGATGTTATCTGGAATACCTTACTGAGTTCTGACGTTTGAATATATGTCGGA	1200
DB	1141	CAGATTGATGTTATCTGGAATACCTTACTGAGTTCTGACGTTTGAATATATGTCGGA	1200
QY	1201	ATGAAGAGATGCTCTTCCAGGTTGATATGTTCCGTTGGTGAAGTTTCACAAGGTTGTTG	1260
DB	1201	ATGAAGAGATGCTCTTCCAGGTTGATATGTTCCGTTGGTGAAGTTTCACAAGGTTGTTG	1260
QY	1261	GATGCTACTGCAAGTTGCTCAGAGAGATACATCAAACTGCAAGTACCAACATCTCG	1320
DB	1261	GATGCTACTGCAAGTTGCTCAGAGAGATACATCAAACTGCAAGTACCAACATCTCG	1320
QY	1321	CAGAAAGAAAGCAAGATGCAAGTTAATTTGAAGTGAATTAAGACTTTTACAGAGATG	1380
DB	1321	CAGAAAGAAAGCAAGATGCAAGTTAATTTGAAGTGAATTAAGACTTTTACAGAGATG	1380
QY	1381	TATGAGCCTTAAGTGGTGTCCACAACCTTAACATCAGGTTGGAAGTGAATTAAGTGT	1440
DB	1381	TATGAGCCTTAAGTGGTGTTCACAACCTTAACATCAGGTTGGAAGTGAATTAAGTGT	1440
QY	1441	TTTGAAGGATGCTACTTCAACTACCTGATGTTGAAGTTGAACAATCTGGAAGAACGGTAA	1500
DB	1441	TTTGAAGGATGCTACTTCAACTACCTGATGTTGAAGTTGAACAATCTGGAAGAACGGTAA	1500
QY	1501	TATGGTCCCTTGAACCTTACTTTTGGGTAACTGGAACAAGTTGATCAAGGCCAAATGG	1560
DB	1501	TATGGTCCCTTGAACCTTACTTTTGGGTAACTGGAACAAGTTGATCAAGGCCAAATGG	1560
QY	1561	TTGTGGAGATCCCTTAAGAGATCTTCAAAAACAAGTATATCCCTTAACCTCTTAAG	1620
DB	1561	TTGTGGAGATCCCTTAAGAGATCTTCAAAAACAAGTATATCCCTTAACCTCTTAAG	1620
QY	1621	GAGCTTAAGCAAGACTTAA 1638	

Db 1545 TATGACCCCTATGCGGGGTTCCAGACCCCAACACGAGTGAGGTGTAAGGTGTG 1604
 Qy 1441 TTGAGAGGATGCTACTTCACTCACTCCGATGTTGATGTAACAACGGAAGAGGTAAAG 1500
 Db 1605 TTGAGAGGATGCTACTTCACTCACTCCGATGTTGATGTAACAACGGAAGAGGTAAAG 1664
 Qy 1501 TATGAGCTCTGGAACTTACTTTTGGGTAACTGAACGATGATGTAAGGCAAAATG 1560
 Db 1665 TATGAGCTCTGGAACTTACTTTTGGGTAACTGAACGATGATGTAAGGCAAAATG 1724
 Qy 1561 TTGAGGATGCTTACGAGATCTTCAAAACAACTGATCTTCTTAACTTAACTTTAAAG 1620
 Db 1725 TTGAGGATGCTTACGAGATCTTCAAAACAACTGATCTTCTTAACTTAACTTTAAAG 1784
 Qy 1621 GAGCCTAAGCAGACTAAATAGTAG 1644
 Db 1785 GAGCCTAAGCAGACTAAATAGTAG 1808

RESULT 4
 AR159722
 LOCUS AR159722 1801 bp DNA linear PAT 17-OCT-2001
 DEFINITION Sequence 30 from patent US 6251626.
 ACCESSION AR159722
 VERSION AR159722.1 GI:16222490
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 1801)
 Stougaard, P. and Hansen, O. Cal.
 Recombinant hexose oxidase, a method of producing same and use of
 such enzyme
 Patent: US 6251626-A 30 26-JUN-2001;
 Location/Qualifiers
 1..1801
 /organism="unknown"
 BASE COUNT 441 a 448 c 492 g 420 t
 ORIGIN

Query Match 77.8%; Score 1279.2; DB 6; Length 1801;
 Best Local Similarity 86.1%; Pred. No. 0;
 Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

Qy 1 ATGGGCTACTTTGGCAAAAAGACCCAGGTTACATTTGATGACCTCAACGCTGTACT 60
 Db 84 ATGGCTACTCTTCTCAGAAAAGACCCGCTTATTTGATGATGATCAACGCGGCACC 143
 Qy 61 CCAGACAAAGCTGACCCAGATGTCATCATGAAGCAAGGTTTCAACAGAAATGAGATT 120
 Db 144 GGGGACAAAGCCGACCCAGCTCTCCCTCATGACAGAGGCTTCAACGCGCTGGATT 203
 Qy 121 GGTACCAACATGATTTGTTTACGTCGTTTACATCTCAACAAAGTCTTGTACTGCTTTG 180
 Db 204 GGAACCTAATATGATTTGTTTATGTCGTACACTCTCAAGGCTCTGTACTGACATT 263
 Qy 181 GACAGAGCTATGAGAAAGTGTCTCCAGGTACCGTACAGATCGTTTCTGTGTCACGT 240
 Db 264 GACCGGCTATGAGAAAGTGTCTCCGAGTACAGATCGTCTCTGCGGCCCAATTGC 323
 Qy 241 TAGAAGACTTCTGTTTTCAGCAAGATGTCAGAGGCTATATCAACGTTACTGTTGGTT 300
 Db 324 TAGAGAGACTTCTGTTTTCAGCAAGATGTCAGAGGCTATATCAACGTTACTGTTGGTT 383
 Qy 301 GAATCTGTTTACAGCAGAGATAGAGGTTACTTGTCTTTCGCTGACCAACGTTGGGTT 360
 Db 384 GAGAGGTGTTATGACAGCATAGAGGTTACTTGTCTGACAGATGAGATACAAATTTGGGG 443
 Qy 361 TCGTTCAACACCTTGTTCAGAGCAACGATGAGGTTTTCAGAGGTTTCTGTTACTCC 420
 Db 444 TCGTTCAACACCTTGTTCAGAGCAACGATGAGGTTTCTGTTACTCC 503

Qy 421 GTGCGTTGGGTGTGCACATTTGCTGAGAGGTGACGATATTTTGGCCAGATTGACCGT 480
 Db 504 GTGCGCTCGGTGTGCACATTTGCTGAGAGGTGACGATATTTTGGCCAGATTGACCGT 563
 Qy 481 TTGCAAGTGTGATTTGTTATCCGCTGTGTAAGTTGTCTTAAAGCAGTCTTTACCGAAGC 540
 Db 564 CTCGCCGTGATTTGGTCCAGCGCGGTGAGGTGTCTGTATACCACTCTCCACCGAAGC 623
 Qy 541 TCTGTTCTTAAGTATGCTTACAAAGATTTCCGAAGGTTACGAGGTTGTTTGGGCT 600
 Db 624 TCGGTACTTAAGTATGCTTACAAAGATTTCCGAAGGTTACGAGGTTGTTTGGGCT 683
 Qy 601 CACACTGTGAGGTGAGAGTAACTTCTGATTTATCCAAATACTACTTCAAGATTG 660
 Db 684 CACACAGGTGGCGGTGGCGAAACTTTGAGATCATCAACAAATACTACTTCAAGATTG 743
 Qy 661 CCAATGTCTCCAAAGAGGTGATCTGCTTCTTAACTTAACTTCTTGGAGCGGTTCACT 720
 Db 744 CCAATGTCTCCAAAGAGGTGATCTGCTTCTTAACTTAACTTCTTGGAGCGGTTCACT 803
 Qy 721 AAGAGTGCCTTGCAGAGATTTGTTGACTAAGTACTTCAAGTGTGATGATTTGGAG 780
 Db 804 AAGAGTGCCTTGCAGAGATTTGTTGACTAAGTACTTCAAGTGTGATGATTTGGAG 863
 Qy 781 AATAGTGTGATTAAGTTCCAAATCTTCCACCAAGAGCTGAAGAGTTGTATGATTTG 840
 Db 864 AATAGTGTGATTAAGTTCCAAATCTTCCACCAAGAGCTGAAGAGTTGTATGATTTG 923
 Qy 841 TATACATCTTACTCTTACAGAGCGCCGAGAGAAAGTTGCCAAGACAGACATATCTTTG 900
 Db 924 TATACATCTTACTCTTACAGAGCGCCGAGAGAAAGTTGCCAAGAGCTTCAATCTTTG 983
 Qy 901 GAGGCTGACATTTGAACAGATCTTCAAAACATGCGAGCTTCAAAAGCTTGTGCTCAT 960
 Db 984 GAGGCTGACATTTGAACAGATCTTCAAAACATGCGAGCTTCAAAAGCTTGTGCTCAT 1043
 Qy 961 GCTGTTGGGCTCTTCCCTGTATGACCTTGAAGAGACACACATCTTCAAGTCTTAT 1020
 Db 1044 GCTGTTGGGCTCTTCCCTGTATGACCTTGAAGAGACACACATCTTCAAGTCTTAT 1103
 Qy 1021 ATGATGACGAGACTATGAGACTTACCTTTCTTACGCTTTGACTGAGACTATCTTCA 1080
 Db 1104 ATGATGACGAGACTATGAGACTTACCTTTCTTACGCTTTGACTGAGACTATCTTCA 1163
 Qy 1081 GGTCTTAATCAGAGAGGTATGATCAAGTCTGTTATCATGATCAAGAGCTTCCAGCTTC 1140
 Db 1164 GGTCTTAATCAGAGAGGTATGATCAAGTCTGTTATCATGATCAAGAGCTTCCAGCTTC 1223
 Qy 1141 CAGATTGATGTTATCTGAAATATCTTACTGAGGTTCTGACGTTTGAAGTGTGCGGA 1200
 Db 1224 CAGATTGATGTTATCTGAAATATCTTACTGAGGTTCTGACGTTTGAAGTGTGCGGA 1283
 Qy 1201 ATGAAGATGCTCTTCTTCAAGTTGATGATGTTGCTGTGAGATTCAAGGTTGTTGG 1260
 Db 1284 ATGAAGATGCTCTTCTTCAAGTTGATGATGTTGCTGTGAGATTCAAGGTTGTTGG 1343
 Qy 1261 GATGCTACGCAAGTGTCTCAGAGAGTATCATCAATCAACGATGACAGACATCTG 1320
 Db 1344 GATGCTACGCAAGTGTCTCAGAGAGTATCATCAATCAACGATGACAGACATCTG 1403
 Qy 1321 CAGAAAGAGCAAGAGTATGATTAAGTGTGATTTAGAGACTTTTACAGAGAGATG 1380
 Db 1404 CAGAAAGAGCAAGAGTATGATTAAGTGTGATTTAGAGACTTTTACAGAGAGATG 1463
 Qy 1381 TATGAGCTTATGTTGTTGTTTCAAGCCTTAACTCAAGTTGAGAGTGTAAAGTGT 1440
 Db 1464 TATGAGCTTATGTTGTTGTTTCAAGCCTTAACTCAAGTTGAGAGTGTAAAGTGT 1523
 Qy 1441 TTGAGGAGTGTCTACTTCAACATCAAGTGTGAGCTTGAACATCTGAAAGAACGTTAG 1500
 Db 1524 TTGAGGAGTGTCTACTTCAACATCAAGTGTGAGCTTGAACATCTGAAAGAACGTTAG 1583
 Qy 1501 TATGAGCTTATGTTGTTTCAAGCCTTAACTCAAGTTGATCAAGGCTCAAAATG 1560

|||||
Db 1584 TATGTCCTCCCTGAACTTATCTTTTGGTACCTGACCGCCCTCATCAAGGCGCAATGG 1643
Qy 1561 TTGGGGATCTCTACGAGATCTTCAACAAACAGCTATCCCTACTAAACCTTAAAG 1620
Db 1644 TTGGGGATCTCCAAAGAGATCTTCAACAAACAGAGCAATCCCTACTAAACCTTAAAG 1703
Qy 1621 GAGCCTAAGCAGACTAATAGTAG 1644
Db 1704 GAGCCCAAGCAGAGAAATAGTAG 1727

RESULT 5
AX030569 1801 bp DNA linear PAT 20-SEP-2000
LOCUS Sequence 30 from Patent EP1020523.
DEFINITION AX030569
VERSION AX030569.1 GI:10278111
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1801)
AUTHORS Hansen, O.C. and Stougaard, P.
TITLE Recombinant hexose oxidase, a method of producing same and use of such enzyme
JOURNAL Patent: EP 1020523-A 30 19-JUL-2000;
BIOTECHNOLOGISK INST (DK)
FEATURES
source Location/Qualifiers
1..1801
/organism="unidentified"
/db_xref="taxon:32644"
84..1724
/note="unamed protein product"
/codon_start=1
/protein_id="CAC09603.1"
/db_xref="GI:10278112"
CDS
/translation="MATLPQKDPGYIVIVNAGTADKPRLPSMKQGFNRMIQTNI
DFYVYVTPQACTALDRAMEKSPGTVRLVSGGH CYEDPEVCYALINVTGLVES
GYDDDRGVFVSGDTNMGSPFKTLFRDHGRVLPQSSCYVGLGGHIVGGGDLIARLHG
LPVDMLSGVVYVVKPVLTEDSVLYKYHKDEGNDGELFPAHHTGGGSPBIIITKYFK
DLPSRPGVIVASNLHPSWDETRDALDDLTFTKFLKARCMKNTVGFQIFHDAEEF
VMVLYTSYNDAREVAQDRHYHLEADI EOIYKTCETPKALGGHAGAPPPVPRRH
TSKTSYMHDETMVYPFALTEITINGSPNGRKYKSYMTKDPDFDIDIVIMKYLEV
PDGLTSAEMDALLQVDFGSEIHKVMDATVAVOREYIIXLOQYTMQBEDQAVNL
KMTIRPEEMVEPGLVGPDPNTQVSEKGVFEGCYRNPVVDLNMNKNKYGALLEYF
LGNLRITKAKMLMDPNEIFTNKQSLPTKLEPKQK"

BASE COUNT 441 a 448 c 492 g 420 t
ORIGIN
Query Match 77.8%; Score 1279.2; DB 6; Length 1801;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

Qy 1 ATGGCTACTTTGCCAAGAAAGAGCCAGGTACATTGTTATGACGTCAACGCTGTACT 60
Db 84 ATGGCTACTCTTCTCAGAAAGAGCCCGGTATATGTAATGATGCAACGCGGACCC 143
Qy 61 CCAGACAGCTGACCCCAAGATTGCCATCCATGAAGCAAGTTTCAACAGAAATGATT 120
Db 144 GCGGACAGCGGAGCCACGCTCTCCCTCATGAAGAGGAGGCTTCAACCGCCGCTGATT 203
Qy 121 GGATCCAAATCATGATTTGTTAGTGTGTTTACCTCACAAGGTGCTTGAATGCTTTG 180
Db 204 GGAATCATATCATGATTTGTTATGTGTGATGACCTCTCAAGGTGCTTGAATGCTTT 263
Qy 181 GACAGACTATGAAAAAGTGTTCACAGATACCGTCAGAAATCGTTTGTGTGCTACTGT 240
Db 264 GACCGTCTATGAAAAAGTGTTCACCGATACAGTCAAGATCGTCTGTGCGGCAATTCG 323
Qy 241 TACGAAGACTTGTGTTTGCAGAAATGTGTCAAGGCTATTTCAACGTTACTGTTGGTT 300
Db 324 TACGAGACTTGTGTTTGCAGAAATGTGTCAAGGCTATTTCAACGTTACTGTTGCTTGT 383

Qy 301 GAATCTGTTAAGACAGACAGTAAAGATTACTTGTCTCTTCCGGTGACCAACTGGGGT 360
Db 384 GAGAGGTGTTAAGACAGACAGTAAAGATTACTTGTCTCTTCCGGTGAGATCAAAATGGGGC 443
Qy 361 TCTTCAAGACCTTGTTCAGAGACCAACGATTAAGTTTCCAGGTGTCTCTGTTACTTC 420
Db 444 TCTTTCMAACCTTGTTCAGAGACCAACGATTAAGTTTCCAGGTGTCTCTGTTACTTC 503
Qy 421 GTCCGTTGGGTGTGTCATGTTGCGGTGAGGTGACGATTTTGGCCAGATTGACCGGT 480
Db 504 GTCCGCTCGGTGTGTCATGTTGCGGTGAGGTGACGATTTTGGCCAGATTGACCGGT 563
Qy 481 TTCCAGTGTGATTTGATTCGCGTGTGAAGTTGTCTTGAACCAAGTCTTGAACGAGAC 540
Db 564 CTCCTCCGTGATTTGGGTACGCGCGGTGAGGTGTCTGTTAAGCCAGTCTTCAACGAGAC 623
Qy 541 TCTGTTCTTAAGTACGTTTCAAGAGATTCGGAAGTAAAGAGGTGATGTTTGGGCT 600
Db 624 TCGGTACTCAAGTATGTGACAAAGATTCGGAAGCAACGAGGGAGCTCTTTGGGCA 683
Qy 601 CACACTGTGAGAGGTGAGGTAACTTCGATATTTATCACCAATATCACTTCAAGGATTG 660
Db 684 CACACAGGTGCGGTGCGGAAACTTTGGAATCATACCAATATCACTTCAAGGATTG 743
Qy 661 CCAATGTCTCAAGAGGTGTCAATGCTTCTTAACCTTCACTTCTTGGACGATTACT 720
Db 744 CCAATGTCTCAAGAGGTGTCAATGCTTCTTAACCTTCACTTCTTGGACGATTACT 803
Qy 721 AGAGATGCTTCAAGATTTGTTGACTAATGTTCTTCAAGTGTGATGATTTGAAG 780
Db 804 AGAGATGCTTCAAGATTTGTTGACTAATGTTCTTCAAGTGTGATGATTTGAAG 863
Qy 781 AATAGTGTGATGATTTGATTTCAAAATCTTCCACAGAGAGGTGATGATTTGATTTG 840
Db 864 AATAGTGTGATGATTTGATTTCAAAATCTTCCACAGAGAGGTGATGATTTGATTTG 923
Qy 841 TATACATCTTACTTCAACGACCGGAGAGAGATTTCCCAAGACAGACATATGATTG 900
Db 924 TATACATCTTACTTCAACGACCGGAGAGAGATTTCCCAAGACAGACATATGATTG 983
Qy 901 GAGGCTGACATTGAACAGATCTTCAAAACATGCGACCTTCAAAAGCTTTGGGTGCT 960
Db 984 GAGGCTGACATTGAACAGATCTTCAAAACATGCGACCTTCAAAAGCTTTGGGTGCT 1043
Qy 961 GGTGTTGGGCTCTTCCCTGTTAGACCTTGAAGAGACACATCCAAAGCTTCTTAT 1020
Db 1044 GGTGTTGGGCTCTTCCCTGTTAGACCTTGAAGAGACACATCCAAAGCTTCTTAT 1103
Qy 1021 ATGCATGACGAGACTATGATGATACCTTTCTACGCTTTGATGAGACTATCAACGATTCC 1080
Db 1104 ATGCATGACGAGACTATGATGATACCTTTCTACGCTTTGATGAGACTATCAACGATTCC 1163
Qy 1081 GGTCTTATCAAGAGAGTAAAGTAAAGTGTCTTATATATCAAGGACTTTCAAGACTTC 1140
Db 1164 GGTCTTATCAAGAGAGTAAAGTAAAGTGTCTTATATATCAAGGACTTTCAAGACTTC 1223
Qy 1141 CAGATTGATTTATCGAATAATCCTTACTAGAGTTTCCGAGGTTTGACTAGTGCCGA 1200
Db 1224 CAGATTGATTTATCGAATAATCCTTACTAGAGTTTCCGAGGTTTGACTAGTGCCGA 1283
Qy 1201 ATGAAGATGCTCTTCTTCAAGTTGATGTTCCGTTGAGATTTCAAGGTTGTTGG 1260
Db 1284 ATGAAGATGCTCTTCTTCAAGTTGATGTTCCGTTGAGATTTCAAGGTTGTTGG 1343
Qy 1261 GATGCTTACTGAGTGTCTCAAGAGATGATCATCATCAACTGCAAGTACCAAGATCTGG 1320
Db 1344 GATGCTTACTGAGTGTCTCAAGAGATGATCATCATCAACTGCAAGTACCAAGATCTGG 1403
Qy 1321 CAGGAAGAGACAGAGATGATTAAGTGAATTTAGAGACTTTTGAAGAGGATG 1380
Db 1404 CAGGAAGAGACAGAGATGATTAAGTGAATTTAGAGACTTTTGAAGAGGATG 1463

QY	1381	TATAGCCTTAATGGTGGTTCACAGACCCCTAACTCAGGTTGAGAGGTGAAGGTGTT	1440
Db	1664	TATAGCCGTTATGGCGGGGTTCCAGACCCCAACGACAGGTGGAGAGTGGTAAAGGTGTG	1523
QY	1441	TTTGAGGAGTCTAATCTCAACTACCCCTGATGTTGACTTTGAACAACTGAGAAACGGTAAAG	1500
Db	1524	TTTATGAGGAGTCTCTACTTCAACTACCCCGGATGTGACTTTGAACAACTGGAAACAGGCAAG	1583
QY	1501	TATGGTGCTTGGAACTTACTTTTGGGTAACTCGAACAAGATGATCAAGGCCAAATAG	1560
Db	1584	TATGTGTCCCTCGAACTTACTTTTGGGTAACTGAACGGCCTCATCAAGGCCAAATAG	1643
QY	1561	TTTGTGGATCTTAACGAGATCTTCACAAAACAACAGTCTATCTCTAAACCTCTTAAAG	1620
Db	1644	TTGTGGATCTCCCAACGAGATCTTCACAAAACAACAGAGATCTCTACTAAACCTCTTAAAG	1703
QY	1621	GAGCCTAAGCAGACATAAATAAGTAG	1644
Db	1704	GAGCCCAAGCAGACATAAATAAGTAG	1727

LOCUS	AX155099/c	126 bp	DNA	linear	PAT 22-JUN-2001
DEFINITION	Sequence 17 from Patent WO0138544.				
ACCESSION	AX155099				
VERSION	AX155099.1				
KEYWORDS	GI:14536647				
SOURCE	synthetic construct.				
ORGANISM	synthetic construct.				
REFERENCE	artificial sequences.				
AUTHORS	1 (bases 1 to 126)				
	Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,				
	Poulsen,C.H. and Zargahi,M.R.				
TITLE	Method for purifying proteins				
JOURNAL	Patent: WO 0138544-A 17 31-MAY-2001;				
	DANISCO A/S (DK)				
FEATURES	Location/Qualifiers				
source	1..126				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="Synthetic oligonucleotide"				
BASE COUNT	33 a 35 c 23 g 35 t				
ORIGIN					
Query Match	7.7%; Score 126; DB 6; Length 126;				
Best Local Similarity	100.0%; Pred. No. 1.2e-22;				
Matches 126; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	1210 GCCTTCTTCACGGTTGATATCTTGGGCGGTAGATTCACAGGTTGTTGGGATGCTACT 1269				
DB	126 GCCTTCTTCACGGTTGATATCTTGGGCGGTAGATTCACAGGTTGTTGGGATGCTACT 67				
QY	1270 GCAGTGGCTCAGAGAGATACATCAACTGACGATCCAGACATCTGCGCAGGAGAA 1329				
DB	66 GCAGTGGCTCAGAGAGATACATCAACTGACGATCCAGACATCTGCGCAGGAGAA 7				
QY	1330 GACAAAG 1335				
DB	6 GACAAAG 1				
RESULT 7					
LOCUS	AX155088	120 bp	DNA	linear	PAT 22-JUN-2001
DEFINITION	Sequence 6 from Patent WO0138544.				
ACCESSION	AX155088				
VERSION	AX155088.1				
KEYWORDS	GI:14536636				
SOURCE	synthetic construct.				
ORGANISM	synthetic construct				
REFERENCE	artificial sequences.				
	1 (bases 1 to 120)				

AUTHORS	Johansen,C.L., Kj Rulff,S.R., Madrid,S.M., Pedersen,H., Poulsen,C.H. and Zargahi,M.R.
TITLE	Method for purifying proteins
JOURNAL	Patent: WO 0138544-A 6 31-MAY-2001, DANISCO A/S (DK)
FEATURES	Location/Qualifiers
source	1..120 /organism="synthetic construct" /db_xref="taxon:32630" /note="Synthetic oligonucleotide"
BASE COUNT	26 a 26 c 31 g 37 t
ORIGIN	

Query	Best Local Similarity	7.3%; Score 120; DB 6; Length 120;	Matches 120; Conservative 100.0%; Pred. No. 5.1e-21;	Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	280	ATCAGCTTACTGTTGGTTGTAATCTGCTTTACGACGACGACGATAGAGTTACTTGTCTCT	339	
Db	1	ATCAACCTTACTGTTGGTTGTAATCTGCTTTACGACGACGACGATAGAGTTACTTGTCTCT	60	
Qy	340	TCGGGTACACCAACTGCGGGGTTCCCTTCAGACCTGTTTCAGAGCCACGGTAGAGTTTG	399	
Db	61	TCGGGTACACCAACTGCGGGGTTCCCTTCAGACCTGTTTCAGAGCCACGGTAGAGTTTG	120	
RESULT 8				
LOCUS	AX155097/c	120 bp	DNA	linear
DEFINITION	Sequence 15 from Patent WO0138544.			PAT 22-JUN-2001
ACCESSION	AX155097			
VERSION	AX155097.1			GI:14536645
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Query Match				
Best Local Similarity				
Matches 120; Conservative				
Qy	1027	GACGAGACTATGACCTATCCCTTTCTACGCTTTGACGACGATCAACGGTTCCGGTCT	1086	
Db	120	GACGAGACTATGACCTATCCCTTTCTACGCTTTGACGACGATCAACGGTTCCGGTCT	61	
Qy	1087	AATAGAGAGGTAAAGTACAGTCTGCTTACATGATCAAGACCTTCCAGACTTCCAGATT	1146	
Db	60	AATCAGAGAGGTAAAGTACAGTCTGCTTACATGATCAAGACCTTCCAGACTTCCAGATT	1	
RESULT 9				
LOCUS	AX155092	118 bp	DNA	linear
DEFINITION	Sequence 10 from Patent WO0138544.			PAT 22-JUN-2001
ACCESSION	AX155092			
VERSION	AX155092.1			GI:14536640
KEYWORDS				
SOURCE				
ORGANISM				

REFERENCE 1 (bases 1 to 118)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Poulsen,C.H. and Zargahi,M.R.
JOURNAL Method for purifying proteins
PATENT: WO 0138544-A 10 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1.118
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 30 a 27 c 22 g 39 t
ORIGIN
Query Match 7.2%; Score 118; DB 6; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.8e-20;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 639 CAATACTACTTCAGAGATTGCGCAATGCTCCAGAGGTGTCATCGCTTCTAATTACA 698
DB 1 CAATACTACTTCAGAGATTGCGCAATGCTCCAGAGGTGTCATCGCTTCTAATTACA 60
QY 699 CTCTCTTGGGACGGTTTCACTAGAGATGCTTGCAGATTGTTGACTAAGTCTTC 756
DB 61 CTCTCTTGGGACGGTTTCACTAGAGATGCTTGCAGATTGTTGACTAAGTCTTC 118

RESULT 10
AXIS5102 117 bp DNA linear PAT 22-JUN-2001
LOCUS AXIS5102
DEFINITION Sequence 20 from Patent WO0138544.
ACCESSION AXIS5102
VERSION AXIS5102.1 GI:14536650
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 117)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Poulsen,C.H. and Zargahi,M.R.
JOURNAL Method for purifying proteins
PATENT: WO 0138544-A 20 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1.117
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 36 a 23 c 26 g 32 t
ORIGIN
Query Match 7.1%; Score 117; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.3e-20;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1489 AAGAACGGTAGTAGTGCTTGGAACTTACTTTTGGGTAACTGAAAGATTGATC 1548
DB 1 AAGAACGGTAGTAGTGCTTGGAACTTACTTTTGGGTAACTGAAAGATTGATC 60
QY 1549 AAGCCCAATGCTGTGGGATCTTAACGAGATCTTCAACAACAACAGTCTATCCCT 1605
DB 61 AAGCCCAATGCTGTGGGATCTTAACGAGATCTTCAACAACAACAGTCTATCCCT 117

RESULT 11
AXIS5093/c 118 bp DNA linear PAT 22-JUN-2001
LOCUS AXIS5093
DEFINITION Sequence 11 from Patent WO0138544.
ACCESSION AXIS5093
VERSION AXIS5093.1 GI:14536641
KEYWORDS
SOURCE synthetic construct.
ORGANISM

REFERENCE 1 (bases 1 to 118)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Poulsen,C.H. and Zargahi,M.R.
JOURNAL Method for purifying proteins
PATENT: WO 0138544-A 11 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1.118
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 39 a 26 c 20 g 33 t
ORIGIN
Query Match 7.1%; Score 116.4; DB 6; Length 118;
Best Local Similarity 99.2%; Pred. No. 4.8e-20;
Matches 117; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 732 GCAGATTGTTGACTTAAGTACTTCAAGTTGGCTAGATGTGAGAGAACTGTTGG 791
DB 118 GCAGATTGTTGACTTAAGTACTTCAAGTTGGCTAGATGTGAGAGAACTGTTGG 59
QY 792 TAAGTCCAAATCTTCCACGACGCTGAAGATTGTTATGTACTTATACATCC 849
DB 58 TAAGTCCAAATCTTCCACGACGCTGAAGATTGTTATGTACTTATACATCC 1

RESULT 12
AXIS5091/c 116 bp DNA linear PAT 22-JUN-2001
LOCUS AXIS5091
DEFINITION Sequence 9 from Patent WO0138544.
ACCESSION AXIS5091
VERSION AXIS5091.1 GI:14536639
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 116)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Poulsen,C.H. and Zargahi,M.R.
JOURNAL Method for purifying proteins
PATENT: WO 0138544-A 9 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1.116
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 33 a 31 c 22 g 30 t
ORIGIN
Query Match 7.1%; Score 116; DB 6; Length 116;
Best Local Similarity 100.0%; Pred. No. 6.1e-20;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 547 CTTAAGTACTTCAAGAGATTCGAAAGGTAAAGACGAGTGTGTTGGGCTCACT 606
DB 116 CTTAAGTACTTCAAGAGATTCGAAAGGTAAAGACGAGTGTGTTGGGCTCACT 57
QY 607 GGTGAGGTGAGAGTAACTTGGTATTATACCAATATCTTCAAGATTGGC 662
DB 56 GGTGAGGTGAGAGTAACTTGGTATTATACCAATATCTTCAAGATTGGC 1

RESULT 13
AXIS5101/c 111 bp DNA linear PAT 22-JUN-2001
LOCUS AXIS5101
DEFINITION Sequence 19 from Patent WO0138544.
ACCESSION AXIS5101
VERSION AXIS5101.1 GI:14536649
KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 111)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 19 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1..111
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 30 a 30 c 21 g 30 t
ORIGIN
Query Match 6.8%; Score 111; DB 6; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1399 GTTCCAGACCCCTAACACTGAGGTGAGAGTGAAGGTGTTTGGAGGATGCTACTTC 1458
DB 111 GTTCCAGACCCCTAACACTGAGGTGAGAGTGAAGGTGTTTGGAGGATGCTACTTC 52
QY 1459 AACTACCTGATGTTGACTTGAACAACTGGAAGAACGTAAGTATGTCGCC 1509
DB 51 AACTACCTGATGTTGACTTGAACAACTGGAAGAACGTAAGTATGTCGCC 1
RESULT 14
AXI55089/c 109 bp DNA linear PAT 22-JUN-2001
LOCUS AXI55089
DEFINITION Sequence 7 from Patent WO0138544.
ACCESSION AXI55089
VERSION AXI55089.1 GI:14536637
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 109)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 7 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1..109
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 35 a 38 c 20 g 16 t
ORIGIN
Query Match 6.6%; Score 109; DB 6; Length 109;
Best Local Similarity 100.0%; Pred. No. 4.8e-18;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 375 GTTCCAGAGCCACCGGTAGAGTTTGGCAGGTGTTCTGTACTCCGTCGGTTGGGTGG 434
DB 109 GTTCCAGAGCCACCGGTAGAGTTTGGCAGGTGTTCTGTACTCCGTCGGTTGGGTGG 50
QY 435 TCACATTGCGGTGAGGTGAGCGGTATTTTGGCCAGATTGCACGGTTTG 483
DB 49 TCACATTGCGGTGAGGTGAGCGGTATTTTGGCCAGATTGCACGGTTTG 1
RESULT 15
AXI55090 109 bp DNA linear PAT 22-JUN-2001
LOCUS AXI55090
DEFINITION Sequence 8 from Patent WO0138544.
ACCESSION AXI55090
VERSION AXI55090.1 GI:14536638
KEYWORDS

SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 109)
AUTHORS Johansen,C.L., KJ Rulff,S.R., Madrid,S.M., Pedersen,H.,
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 8 31-MAY-2001;
DANISCO A/S (DK)
FEATURES
source 1..109
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 22 a 23 c 28 g 36 t
ORIGIN
Query Match 6.6%; Score 109; DB 6; Length 109;
Best Local Similarity 100.0%; Pred. No. 4.8e-18;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 462 TTTGGCCAGATTGCACCGTTTCCAGTGCATTATCCGGTGTGAAGTTGTCTTAA 521
DB 1 TTTGGCCAGATTGCACCGTTTCCAGTGCATTATCCGGTGTGAAGTTGTCTTAA 60
QY 522 GCCAGTCTGACCGAAGACTGTCTTAAAGTACGTTCAAGATTCC 570
DB 61 GCCAGTCTGACCGAAGACTGTCTTAAAGTACGTTCAAGATTCC 109

Search completed: June 15, 2003, 04:51:33
Job time : 4312 secs

Qy	1	MATLPQDPCGYIYIDVNACTPKDPRLPSMQGNGRRRIGINIDFVVVVTPOACATL	60
Qy	1	MATLPQDPCGYIYIDVNACTPKDPRLPSMQGNGRRRIGINIDFVVVVTPOACATL	60
Db	1	MATLPQDPCGYIYIDVNACTPKDPRLPSMQGNGRRRIGINIDFVVVVTPOACATL	60
Qy	61	DRAMEKSPGTVARIVSGHCYEDPFVDECYKAIIIVNTGLVBSGYDDDRGYFVSSGDTNWG	120
Db	61	DRAMEKSPGTVARIVSGHCYEDPFVDECYKAIIIVNTGLVBSGYDDDRGYFVSSGDTNWG	120
Qy	121	SFTLTFRDHRRLVPGSSCYSVGLGHIIVGGNGGIIILARLHGLPVDMLSGVEVYVVKVLTLED	180
Qy	121	SFTLTFRDHRRLVPGSSCYSVGLGHIIVGGNGGIIILARLHGLPVDMLSGVEVYVVKVLTLED	180
Db	121	SFTLTFRDHRRLVPGSSCYSVGLGHIIVGGNGGIIILARLHGLPVDMLSGVEVYVVKVLTLED	180
Qy	181	SVLKYYVHDSSEGDGELFMAHTGGGGGNGGIIITTKYYFPDLPMSPRGVIALSNLHFSMDGFT	240
Qy	181	SVLKYYVHDSSEGDGELFMAHTGGGGGNGGIIITTKYYFPDLPMSPRGVIALSNLHFSMDGFT	240
Db	181	SVLKYYVHDSSEGDGELFMAHTGGGGGNGGIIITTKYYFPDLPMSPRGVIALSNLHFSMDGFT	240

QY 241 RDALQDLITKFKLARCDMKNITVGKFOIFHOAAEFVWYLYTSYSDAEREVAQDRHYHL 300
DB 241 RDALQDLITKFKLARCDMKNITVGKFOIFHOAAEFVWYLYTSYSDAEREVAQDRHYHL 300
QY 301 EADIEQIYKTCCEPTKALGSHAGMAPPPVRPRKHTSKTSYMHDETMDFPYALTETINGS 360
DB 301 EADIEQIYKTCCEPTKALGSHAGMAPPPVRPRKHTSKTSYMHDETMDFPYALTETINGS 360
QY 361 GPNORGKYSAYMIKDPDPQIDIVIMKYLTEVPDGLTSAEMKDALLOVDMFGSEIHKVM 420
DB 361 GPNORGKYSAYMIKDPDPQIDIVIMKYLTEVPDGLTSAEMKDALLOVDMFGSEIHKVM 420
QY 421 DATAVAGREYIIKLQYQTYWQEDKDAVNLKIMRDFEEMYPEYGVDPDPTQVESGKV 480
DB 421 DATAVAGREYIIKLQYQTYWQEDKDAVNLKIMRDFEEMYPEYGVDPDPTQVESGKV 480
QY 481 PEGCYFNPVDVNLNWKNGKYGALFLYGLNLRILKAKWIMDNEIFTNKOSIPTKPLK 540
DB 481 PEGCYFNPVDVNLNWKNGKYGALFLYGLNLRILKAKWIMDNEIFTNKOSIPTKPLK 540
QY 541 EPKOTK 546
DB 541 EPKOTK 546

RESULT 2

US-09-824-053-31
Sequence 31, Application US/09824053
Patent No. US20020106725A1

GENERAL INFORMATION:

APPLICANT: Peter Stougaard
Ole Cai Hansen

TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME

NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESS: Hutton & Williams
STREET: 1900 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006-1109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/824.053
FILING DATE: 03-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/669,304
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Stanislaus Akeman

REGISTRATION NUMBER: 28,562
REFERENCE/DOCKET NUMBER: <Unknown>

TELEPHONE: (202) 955-1926
TELEFAX: (202) 778-2201
TELEX: NO. US20020106725A1e

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid

MOLECULAR TYPE: linear
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-824-053-31

Query Match 99.7%; Score 2989; DB 10; Length 546;
Best Local Similarity 99.8%; Pred. No. 1,9e-261;

Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MATLPQKDPQYIYIDVNAAGTADKPPDLPSMKGFNRMRGTNIDFYVYVYTGACTAL 60
DB 1 MATLPQKDPQYIYIDVNAAGTADKPPDLPSMKGFNRMRGTNIDFYVYVYTGACTAL 60
QY 61 DRAMEKSPETVIVSGHGYEDFVDECYKAIINTGLVESYDDDRGVFSSGDTMWG 120
DB 61 DRAMEKSPETVIVSGHGYEDFVDECYKAIINTGLVESYDDDRGVFSSGDTMWG 120
QY 121 SFTLFRDHRVLPGSCSVSGHIGVGGDGLLARHGLPVDMLSGVGVVVKPVLTED 180
DB 121 SFTLFRDHRVLPGSCSVSGHIGVGGDGLLARHGLPVDMLSGVGVVVKPVLTED 180
QY 181 SVLKVYHKSSEGDGELFMHTGGGNGFIIITKYYFKDLPMSPRGVIASNLHFSMDGFT 240
DB 181 SVLKVYHKSSEGDGELFMHTGGGNGFIIITKYYFKDLPMSPRGVIASNLHFSMDGFT 240
QY 241 RDALQDLITKFKLARCDMKNITVGKFOIFHOAAEFVWYLYTSYSDAEREVAQDRHYHL 300
DB 241 RDALQDLITKFKLARCDMKNITVGKFOIFHOAAEFVWYLYTSYSDAEREVAQDRHYHL 300
QY 301 EADIEQIYKTCCEPTKALGSHAGMAPPPVRPRKHTSKTSYMHDETMDFPYALTETINGS 360
DB 301 EADIEQIYKTCCEPTKALGSHAGMAPPPVRPRKHTSKTSYMHDETMDFPYALTETINGS 360
QY 361 GPNORGKYSAYMIKDPDPQIDIVIMKYLTEVPDGLTSAEMKDALLOVDMFGSEIHKVM 420
DB 361 GPNORGKYSAYMIKDPDPQIDIVIMKYLTEVPDGLTSAEMKDALLOVDMFGSEIHKVM 420
QY 421 DATAVAGREYIIKLQYQTYWQEDKDAVNLKIMRDFEEMYPEYGVDPDPTQVESGKV 480
DB 421 DATAVAGREYIIKLQYQTYWQEDKDAVNLKIMRDFEEMYPEYGVDPDPTQVESGKV 480
QY 481 PEGCYFNPVDVNLNWKNGKYGALFLYGLNLRILKAKWIMDNEIFTNKOSIPTKPLK 540
DB 481 PEGCYFNPVDVNLNWKNGKYGALFLYGLNLRILKAKWIMDNEIFTNKOSIPTKPLK 540
QY 541 EPKOTK 546
DB 541 EPKOTK 546

RESULT 3

US-09-258-031B-75
Sequence 75, Application US/09258031B
Patent No. US20020168735A1

GENERAL INFORMATION:

APPLICANT: STUYVER, Maarten Hendrik

APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor

APPLICANT: SELA-BURLAGE, Marianne Beatrice

APPLICANT: MELCHERS, Leo Stoerd

APPLICANT: VAN DEVENTER-TROOST, Johanna Pietermeelia

APPLICANT: LAGWEG, Wessel

APPLICANT: PONSSTEIN, Anne Silene

APPLICANT: LAGWEG, Wessel

APPLICANT: PONSSTEIN, Anne Silene

TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING

TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING

NUMBER OF SEQUENCES: 75

CORRESPONDENCE ADDRESS:

ADDRESSEE: LADAS & PARRY

STREET: 26 WEST 61 STREET

CITY: NEW YORK

STATE: NY

COUNTRY: USA

ZIP: 10023 - 7604

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.25" Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: WINDOWS 95

SOFTWARE: WORDPERFECT 8

```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/258,031B
? FILING DATE: 25-FEB-1999
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/EP97/04923
? FILING DATE: 04-SEP-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: EP97200831.2
? FILING DATE: 19-MAR-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: EP96202466.7
? FILING DATE: 04-SEP-1996
? INFORMATION FOR SEQ ID NO: 75:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 509 amino acids
? type: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? IS-09-258-031B-75

```

```

Query Match          9.8%; Score 292.5; DB 9; Length 509;
Best Local Similarity 24.9%; Pred. No. 6.4e-18;
Matches 125; Conservative 68; Mismatches 197; Indels 113; Gaps 23

QY      73 RIVSGGCHYEYEDFVDECVK--AIIINVGLAESGYD-DDRGYFVSSGGDTNMGSEFKTLPRD 128
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      82 RVASGGHDFELSLVSNRIEKPFILLDLDSKLQINWDIESNANWQPGAT-IGELLYRTAE 140
QY      129 HGRV-LPGGSCYSVGLAGHIVGGDGLIARHGLPVDMLSGVEVYVKKPVLTEDSVLKYV 186
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      141 KSKIHGFPAGICTSVIGIYVMTGGGVLTKMRKYGLAGDNVLDVKKV-----DANGKLL 193
QY      187 HKHSEGNDELPMANTGGGNGNFGIITKYTKDLPMSRGVYASNLHFSWDFTRDALOD 246
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      194 DRAMGED-LFMAIRGGGASFGIYLAWKIKLVFV-PKTVTVFTV-----TKTLEOD 243
QY      247 LITKYFLIARCDIMKNYCKE-QIFHOAAEFP-VNVLVTSYSDAREVAODRHYLLEADI 304
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      244 ARLK-----TISKWQIISKIIIEIHIRVILRAAGDGNKIVTMYTGLGQFLGK 292
QY      305 EOIYKTCF--PTKAL-----GSHAGMAEPFVPRKRNHTSKTSYMD 344
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      293 GTLLKWEKAPREHIGLQKQDCTEMSWTEALFHHGFPFGSPLEILLQKSLPGKQYF-- 349
QY      345 TMDYFAYALTETTINGSGENQGRKYSAYMIKDFPDQIDIVKXVITEVPDGLTSAEMKDA 404
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      350 -----KATSDPEVKEPPIVIGFKGJIFKRLIE---GNVT----- 378
QY      405 LLOVDMGEGELHKKVMDATAVAQAE-YIIKIQYQYVQWEEBK-DAVNLKMTIDPFEEWYE 462
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      379 FLNMTPRYGMMMSKIPESAIIPPHRNGLTKLLYYANWLENDKTSRKNIMKEIYNM-A 437
QY      463 PYGGVPRPNNQVESGKGQVFEBCGYFNPDPVDL-NMMKNKGKYGALDEL-----YPLGNLRL 515
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      438 PY-----VSSNP---RQAIYVNRDLDPGQNNKNAKVNFTIAKIVGPKYFKGNFDRL 485
QY      516 IKAKMLDPNEIFTNKOSIPTKP 538
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      486 VKIKTKVDPENFFRHQSIIPPMP 508

RESULT 4
US-09-258-031B-73
; Sequence 73, Application US/09258031B
; Patent No. US20020168735A1
; GENERAL INFORMATION:
; APPLICANT: STUIVER, Maarten Hendrik
; APPLICANT: CUSTERS, Jerome Hubertina Henricus Victor
; APPLICANT: SELA-BURLAGE, Marianne Beatrix
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: VAN DEVENTER-TROOST, Johanna Pietermella
; APPLICANT: LAGEWEG, Wessel

```

```

1  APPLICANT: PONSTEIN, Anne Silene
2  APPLICANT: LAGMEG, Wesel
3  APPLICANT: PONSTEIN, Anne Silene
4  TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
5  TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
6  TITLE OF INVENTION: SAME.
7  NUMBER OF SEQUENCES: 75
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: LADAS & PARRY
10 STREET: 26 WEST 61 STREET
11 CITY: NEW YORK
12 STATE: NY
13 COUNTRY: USA
14 ZIP: 10023 - 7604
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: 3.25" Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: WINDOWS 95
19 SOFTWARE: WORDPERFECT 8
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/09/258,031B
22 FILING DATE: 25-FEB-1999
23 CLASSIFICATION: 435
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: PCT/EP97/04923
26 FILING DATE: 04-SEP-1997
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: EP97200831.2
29 FILING DATE: 19-MAR-1997
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: EP96202466.7
32 FILING DATE: 04-SEP-1996
33 INFORMATION FOR SEQ ID NO: 73:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 508 amino acids
36 TYPE: amino acid
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39 OS-09-258-031B-73

```

```

Query Match      9.28: Score 277; DB 9; Length 508;
Best Local Similarity 23.64; Pred. No. 1.6e-16;
Matches 120; Conservative 67; Mismatches 199; Indels 122; Gaps 20;

QY      72  VRIVSGGCHCYED-----FVPEDCVKAIIIV-----TGLVSEGYDDRCGYFWSSG 115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      82  MKIIRSGGHDYDGLSYVTYSGKPFVFLDMFMLRSYDVVDVASKTAVQTCALIGEYVYI- 139

QY      116 DTWVGSKFTLFRDHGRVLPGGSCSVGLGGHHYGGSGGILLARHGLVMDLSGVEVYVKP 175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      140 ---WESKTL-----AYPAGICTVVGHHISGGSGGNMRRKGLTVDNTIDARMY--- 187

QY      176 VLTEDSVLTKVHKHSEGNDELFWAHNGGGGNGFGLITTKYFYKDLPMSPRGVIAINSNHS 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      188 ----DVNGKILDRKMGED--LYWAINGGGGSGYGVVLAVKI- NLVEVPEVTVFTRISRT 240

QY      236 WDEGTRALODLTKTKYKLCRCMKNTVGF--QIFQAAAEFPWIKYTSYNSDAEEVA 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      241 LEONATDIH-----RQOYVAPKLPDELFRIVIDVNGTVSSQKTRVTFIA 288

QY      294 Q-----DRHYHLEADIEQIKYTCSEPTALGHHGMAFPVPYPRKRTSKTSYM 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      289 MFLQDTTLLSIINRR---PELGLVNSDCTETSIQSVLEFWNIQVGSS- 336

QY      342 HDETMDFPYALJETTINGSGPNQRGKYKSAVMIKDPDFQIDVIWKYLTVEVPDGLTSAEM 401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      337 -----TLLLRNQPNVYVYLKRRKSYVREPIRSTGIESLTKMKM-----IEL 375

QY      402 KDALLQVDMEGGEHKKVVMDATVAQRE-VYIKQYQTYQOEEDKDAVNLKVIDPYEEM 460
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      376 EIPMANPYPGEGEGRISSTVTPRPYAGNLKMKIQYQANMRDELITDRVWELTRKLYQFM 435

QY      461 YEPYGVDPNPQTVESGKVPBEGCYFNYPDVDLN-NMKNGK-----YGALELYFLG 510

```

Db 436 -TFVVS-KNPROS-----FVNYRDVDLGINSHNGKISSYVEGKEYG--KKYFAG 480
QY 511 NUNRLIKAKMLDPNEIFTNKOSIPTRP 538
Db 481 NFERLVKIKTRVDSGNFFRNDSIPVLV 508

RESULT 5

US-09-258-031B-16
Sequence 16, Application US/09258031B
Patent No. US2002016735A1
GENERAL INFORMATION:
APPLICANT: STUIVER, Maarten Hendrik
APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor
APPLICANT: SELA-BURLAGE, Marianne Beatrix
APPLICANT: MELCHERS, Leo Sjoerd
APPLICANT: VAN DEVENTER-TROOST, Johanna Pietermella
APPLICANT: LAGMEG, Wessel
APPLICANT: PONSSTEIN, Anne Silene
APPLICANT: PONSSTEIN, Anne Silene
TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
TITLE OF INVENTION: SAME.
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESSES:
ADDRESSEE: LADAS & PARRY
STREET: 26 WEST 61 STREET
CITY: NEW YORK,
STATE: NY
COUNTRY: USA
ZIP: 10023 - 7604
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.25" Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/258, 031B
FILING DATE: 25-FEB-1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP97/04923
FILING DATE: 04-SEP-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP97200831.2
FILING DATE: 19-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP96202466.7
FILING DATE: 04-SEP-1996
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-258-031B-16

Query Match 9.2%, Score 275, DB 9, Length 529,
Best Local Similarity 23.9%, Pred. No. 2.6e-16,
Matches 144, Conservative 64, Mismatches 197, Indels 198, Gaps 26;

QY 22 DKPP-----RLPSMKOGF--NRWIGNIDVVVVYVPG-----AC 57
Db 34 DRAPSPPTIGEVYTPGNSFPVLYONIRKLRNETTTPKPELITAEHVSHIQAAVVC 93
QY 58 TALDRAMKSCPGTVRIVSGHCYED-----FVPECKVAILNV-----TGLIVE 101
Db 94 GKQNRLLIKTR-----SGGHDIEGLSYLNTNQPFYIVMFLRLSINDIDEBETAMVQ 146
QY 102 SGYDDDRGYFVVSGBDTMKGSKFLTFRDHGRVLPGSCYCVGLGHHIVGGGDIARLHGL 161

Db 147 AGATLGEVYVYRIAEKN-----KHG--FPAGVCPITVGGHSGGSGYGLMKRYGL 195
QY 162 PVDMLSGVEVYVAVPVLTEDSVLKYVAKDSGNDGELFMHTGGGCGNFGIITKYYPKDL 221
Db 196 SVNDIYDAQIT-----DYNGLDLRKSMBE--LFWAITGGGVSFGVLAAYKIK-LV 245
QY 222 MSPRGVIAASNLHFSWMDGFTFDALQDLITKYFKIARCDMKNTYVGFQI FHOAAEEFVWLY 281
Db 246 RVEEVTV-----FTIEREEQN----- 263
QY 282 TSYSDAER--EVAQ--DRHYHLEADIEQIYKTCETPKLGLGHAGAPPVVRKRKHTSK 337
Db 264 --LSTJAEVWQVADKLDRDLFLRMTFSVINDT-----NGKTVRAIRP----- 305
QY 338 TSYWDE-----TWDPFFYALTE-----TINGSGNOR-- 365
Db 306 TLVLGSRNLVTLNNDPFLGLQESDCJTEMSVSVLYTGPSPGPTTALLSRTPQL 365
QY 366 --GKYSAVWIKDPDFQIDVWKVLTVPDGLTSAEMKDALLQVDFGGEIHKVYWDAT 423
Db 366 NPFKISDYQNPISKRQEFIFERL-----KELENOMLAFNPYGGRMSEISEFAK 416
QY 424 AVAQRE-YIIKQOYQYVQEDKDAN--LKWTRDYEEMVEPYGVDPNQVESGKV 480
Db 417 PPHRSGNIAKIQYEVNWEDELSDAENRYLFTRLMYDVN-----TFVSKNPRK-- 466
QY 481 FEGCYFNPVDLNNKNGKYGALIEYFLG-----NUNRLIKAKMLDPNEIFTNKOS 533
Db 467 ---AFIANYRDLDLGINSHGRNATBEGVYGHKFKETNYKRLVSVTKYDPPDFNDS 523
QY 534 IPT 536
Db 524 IPT 526

RESULT 6

US-09-258-031B-20
Sequence 20, Application US/09258031B
Patent No. US20020168735A1
GENERAL INFORMATION:
APPLICANT: STUIVER, Maarten Hendrik
APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor
APPLICANT: SELA-BURLAGE, Marianne Beatrix
APPLICANT: MELCHERS, Leo Sjoerd
APPLICANT: VAN DEVENTER-TROOST, Johanna Pietermella
APPLICANT: LAGMEG, Wessel
APPLICANT: PONSSTEIN, Anne Silene
APPLICANT: LAGMEG, Wessel
APPLICANT: PONSSTEIN, Anne Silene
TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
TITLE OF INVENTION: SAME.
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESSES:
ADDRESSEE: LADAS & PARRY
STREET: 26 WEST 61 STREET
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10023 - 7604
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.25" Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/258, 031B
FILING DATE: 25-FEB-1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP97/04923
FILING DATE: 04-SEP-1997
PRIOR APPLICATION DATA:

Db 373 LNPFKISDYKNTISKGFESIFERMKELBNQMLAFNPYGMRMSEISEPAKDFPHRSGN 432
 QY 431 IITLQVOTYQOEKQAVN--LKMIRDFEEMTEPYGVDPDPYQVSGGVGECGFENY 488
 Db 433 IAKIQEVNDEKGVENANRYLNTFRVYDYMTEPFVSKNP-----REAVLNTY 479
 QY 489 PDVILNMKMGK-----YGALELYFL-GNLNRLIKAKMLDPNEIFTNKS IPT 536
 Db 480 RDDDIGVNSHGKAKYVGGWYTG--HKYFKETNYRLIMVTRDPSPNFFRNEQS IPT 534

RESULT 8 US-09-258-031B-71

; Sequence 71, Application US/09258031B
 ; Patent No. US20020168735A1
 ; GENERAL INFORMATION:
 ; APPLICANT: STUIVER, Maarten Hendrik
 ; APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor
 ; APPLICANT: SELA-BURLAGE, Marianne Beatrix
 ; APPLICANT: MELCHERS, Leo Sjoerd
 ; APPLICANT: VAN DEVENTER-TROOST, Johanna Pietermella
 ; APPLICANT: LAGEMEG, Wessel
 ; APPLICANT: PONSSTEIN, Anne Silene
 ; APPLICANT: LAGEMEG, Wessel
 ; APPLICANT: PONSSTEIN, Anne Silene
 ; TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
 ; TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
 ; NUMBER OF INVENTION: SAME.
 ; NUMBER OF SEQUENCES: 75
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LADAS & PARRY
 ; STREET: 26 WEST 61 STREET
 ; CITY: NEW YORK
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10023 - 7604
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.25" Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: WINDOWS 95
 ; SOFTWARE: WORDPERFECT 8
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/258, 031B
 ; FILING DATE: 25-FEB-1999
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/EP97/04923
 ; FILING DATE: 04-SEP-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP97200831.2
 ; FILING DATE: 19-MAR-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP96202466.7
 ; FILING DATE: 04-SEP-1996
 ; INFORMATION FOR SEQ ID NO: 71:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 508 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-258-031B-71

Query Match 8.8%; Score 263; DB 9; Length 508;
 Best Local Similarity 23.2%; Pred. No. 3e-15;

Matches 118; Conservative 67; Mismatches 201; Indels 122; Gaps 20;

QY 72 VRIVSGGCHCED-----FVPEECVAKIINV-----TGLVESGYDDDRGCVFVSSG 115
 Db 82 MKIRSGGHDYDGLSVYVSGKPFVLDMFNLRSDVDVAVASTAVVGTGAILGVEYYII-- 139
 QY 116 DTNMGSEKTLFRDHGRVLPGGSCYVGLGHI VGGGDILARLHGLPDVIMLGVGVVVKP 175

Db 140 ---WEKSKTL-----AYPAGICPTVGVGHI SGGSGYGMWRKRYGLTVDNTIDARNV--- 187
 QY 176 VLTEGSLVYKVKHDSRGNDGELFWAHTGGGNGFIITKYFKDLPMSPRGVIASTMHS 235
 Db 188 ----DVNGKILDRKLMGED--LYMAINGGGGSGYVLLAKY--NLVEVENTVPIISRT 240
 QY 236 WDFETRDALODLLTKYFKLARCDKNTVGKF--QIFHOAAEEFWMLYYSYNDAREVA 293
 Db 241 LEQNATDIIH-----RMQOVAPKLDDELFRVIVDVNVTGSOKVTRTFIA 288
 QY 294 Q-----DRHYHLEADIEQYKCEPTKVLGSHAGNAPFVRPRKHTSKTSYV 341
 Db 289 MELGDTTLLSLINRF---PEGLVRSPTCTETSWIQSVLFWTNIQVSSG----- 336
 QY 342 HDFTMDYPFVALTETINGSQPNRGKSAVMKDFPDQIDVIMKYLTEVPDGLTSAM 401
 Db 337 -----TLLQRNQPVVYLKRSDDYREPISRTGLSIEWKK-----IEL 375
 QY 402 KDALLQVDMFGGHIHKVWDATVAQRE-YIYKLOQTYQOEEDKQAVLKMIRDFEEM 460
 Db 376 EIPYMAFNPGGEMGRISLRVTFPPYRAGNLWKIQGANWRDETLTDRYMELTRKLYQFM 435
 QY 461 YEPYGVPPDPNTOVESGKGVFEGCVFNYPVDLN--WKXNGK-----YGALELYFG 510
 Db 436 -TFVVS-KNPROS-----FPNNDVDVLGINSHGKISSYVEGKRYG--KTYFAG 480
 QY 511 NLNRLIKAKMLDPNEIFTNKS IPTK 538
 Db 481 NFERLVKIKTRVDSGNFFRNEHSI PVL 508

RESULT 9

US-09-932-923-1
 ; Sequence 1, Application US/09932923
 ; Patent No. US20020064577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SOE, JORN B.
 ; APPLICANT: POLSEN, CHARLOTTE H.
 ; APPLICANT: HOSTRUP, PERNILLE B.
 ; TITLE OF INVENTION: A METHOD OF IMPROVING THE PROPERTIES OF A FLOUR DOUGH,
 ; TITLE OF INVENTION: A FLOUR DOUGH IMPROVING COMPOSITION AND IMPROVED FOOD
 ; TITLE OF INVENTION: PRODUCTS
 ; FILE REFERENCE: 59191-000003
 ; CURRENT APPLICATION NUMBER: US/09/932, 923
 ; CURRENT FILING DATE: 2001-08-21
 ; PRIOR APPLICATION NUMBER: 08/676, 186
 ; PRIOR FILING DATE: 1996-07-12
 ; PRIOR APPLICATION NUMBER: 08/483, 870
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: PCT/DK96/00239
 ; PRIOR FILING DATE: 1996-06-04
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 19
 ; TYPE: PRT
 ; ORGANISM: Chondrus Crispus
 ; FEATURE:
 ; OTHER INFORMATION: Description of Sequence: peptide
 ; US-09-932-923-1

Query Match 3.6%; Score 107; DB 10; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.0033;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 DPGYIVDVNAGTPDKXP 26
 Db 1 DPGYIVDVNAGTPDKXP 19

RESULT 10
 US-09-880-748-1466
 ; Sequence 1466, Application US/09880748

Publication No. US20030059937A1
GENERAL INFORMATION:
APPLICANT: Ribben et al.
TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
FILE REFERENCE: PF523
CURRENT APPLICATION NUMBER: US/09/880, 748
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/212, 210
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 60/240, 816
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/276, 248
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/277, 379
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/293, 499
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 3239
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1466
LENGTH: 254
TYPE: PRT
ORGANISM: Homo sapiens
US-09-880-748-1466

Query Match 3.2%; Score 96.5; DB 9; Length 254;
Best Local Similarity 21.2%; Pred. No. 1.2;
Matches 44; Conservative 22; Mismatches 65; Indels 77; Gaps 6;

QY 17 NAGTPDKPDPRLPSMKQGNRRNIGTNIDFVVVYTPQGACTALDRAMKCSPTVRIYS 76
DB 57 NAGYAEKFEGRV-SMTTNI PKK-----TVYNELSLTSEDTAVVYCAAGT----- 100
QY 77 GGHCVDFYFDECVKAIINVTGLVESGYDDDRGYFPVSSGDTWMSKTLFRDHGRVLPBG 136
DB 101 -----GYDILGTGYMGSAFDQWR-----G 120
QY 137 SCYVGLGHIIVGGDGIILARLHLPVDMLSGVEVVVKPVLTEDSVLKVHKDSGNDSG 196
DB 121 TLVTYSSGGGSGSGG-----SSGGSGASQSVLTP-----PSASGTPQ 159
QY 197 LFWAHTGGGGNGFIITKXYFKDLPMS 224
DB 160 TVTISCGGGTNIGRDRTVWYQQVPGTP 187

RESULT 11
US-10-124-880-4
Sequence 4, Application US/10124880
Publication No. US20030026810A1
GENERAL INFORMATION:
APPLICANT: Jorgensen, Per Lina
APPLICANT: Schmoor, Kirk
APPLICANT: Andersen, Lene No. US20030026810A1boc
APPLICANT: Schulte, Martin
APPLICANT: Outtrup, Helle
TITLE OF INVENTION: No. US20030026810A1el Rhamnogalacturonan Hydrolases
FILE REFERENCE: 5572.204-US
CURRENT APPLICATION NUMBER: US/10/124, 880
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US/09/311, 626B
PRIOR FILING DATE: 1999-05-13
PRIOR APPLICATION NUMBER: 0608/98
PRIOR FILING DATE: 1998-05-01
PRIOR APPLICATION NUMBER: 60/084, 358
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FaestSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 622
TYPE: PRT
ORGANISM: Bacillus licheniformis
US-10-124-880-4

Query Match 3.2%; Score 95; DB 9; Length 622;
Best Local Similarity 20.4%; Pred. No. 5.9;
Matches 114; Conservative 81; Mismatches 212; Indels 152; Gaps 33;

QY 31 MKQGNRRN-----IGNIDFVVV--YTPQ---GACTALDRAMKCSPTVRIYSGGHC 80
DB 1 MKKG-KRWKNLLAASLLITLVTFSPQAEADGRTAQAQRMESLNGLVAVKQNGV 59
QY 81 YEDF--VPDECVKAIINV-----TGLVESGYDDDRGYFPVSSGDTWMSKTLFRDHGRVLP 134
DB 60 FVSWRLGTSPSSVSLNVRNGKRLNGSPITSTSTNQDAGDLN----- 103
QY 135 GGSVCYVGLGHIIVGGD-----GILRLH-GLPVDMLSGVEVVVKPVLTEDSVLKVY 186
DB 104 --AVYVRA--VLNREQAPESESVLWQYKSVLPQPAQK-----TPDGV-STI 150
QY 187 HKDSEGNDELFWAHTGGGGNGFIITKXYFKDLPMSPRGVIASNLHFSWDEFTRDALOD 246
DB 151 YSANDASLGDLV-----GDAQYEIPLKV-----DPSNSKDNQDGYTGDLVID 193
QY 247 LTKYFKL-ARCDWKATVKGFOI FHOAAEFVWYLYTYSINDAREVADRHVLEADIE 305
DB 194 A---YKLDGTWMMWRINLGR--NIRAGAHYTOFLVYDFDGDGKAEIAM----- 235
QY 306 QYKTCSEPTK-ALGHAAGWAPPVR-PRKRHTSKTSYM-----HDEIMDYP----- 349
DB 236 ---KTADGTHKDGKAKYIGANADYRNAQRIISGPEYLVFVKDDTGAELITVYEPARGN 292
QY 350 -----FYALTETINGSQPN--ORGYKSAWMIK-DPPQIDIVIMKYLTE 391
DB 293 VADMGDSYGNVRDRFLAGAVYLDGERPSFVMARGVYTRVLVAVNPRGGKLTGLWTFDSD 352
QY 392 VP-DGLTSAEMDALLQVMFGEIHKVWDATAVAQREYIIKLQYQYTWQEBD-KDAYN 449
DB 353 APNGAYIAQGNHSLSVADVDGKDEIYIGAMAV---DHDGGLVSTQMGHDAMHTGN 409
QY 450 LKMIK-----DFEYEMVPPYG-GVPDPNT-----QVESKGVEGECFNYPPVD----- 492
DB 410 LDBSRPGLFVFOVHEHNSNPGLSFRDAKTGKIIMGVHAGKDVGRM---AADIDRYEG 466
QY 493 LNMWKNKYGALEYFLGN 511
DB 467 AEVMAAGSLYTKAGVKIGN 485

RESULT 12
US-09-815-242-12135
Sequence 12135, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191, 078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206, 848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625

;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 1410
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12135
;; LENGTH: 657
;; TYPE: PRT
;; ORGANISM: *Staphylococcus aureus*
US-09-815-242-12135

Query Match 3.2%; Score 95; DB 10; Length 657;
Best Local Similarity 19.1%; Pred. No. 6.4;
Matches 79; Conservative 51; Mismatches 139; Indels 144; Gaps 21;

QY 209 FGIITKXYKFDLPMSPRGVASNLHFSWDFTRDALODLLTKYFKLARCWMKNTVG--- 264
DB 6 FYITTPITY-----PSG-----NLHIGHAYST--VAGDVIAKYKMGQYDVRVYLTGDEH 53
QY 265 --KFOIFHOAA-----EEFV-----MYLYSYNDAREVAQODRHYLEADIEQIYK 309
DB 54 GOKIOEKAOQAKTELEYLDEMIAIGIKQWAKLEISNDDEFRTTEERHKIV--VEQVFE 110
QY 310 TC--EPTKALGAGWAPFVPRPKRHTSKTSYVHDETMQDPFYALTEITNGSGPNQRK 367
DB 111 RLKQSDIYIGEHGWSVP-----DETYYTESLVP--POYENKIIIGSGSPDSCHE 161
QY 368 YKSAWMIKDPDF-----QIDVIMKYLTVEVPDGLTSAEMKDALLQ-----VDM----- 410
DB 162 VE--LVKESYFPNISKYTDRLLEFYDQNDPFIQPSRKKNMNNFIKGLADLAVSRT 218
QY 411 -FGEIHF-----KYVMDATNAQREYIIKLOQYQYWOEDMDANLKMIDFY----- 457
DB 219 SFMWGVHVPSPNKHVVYVWIDALV--NYISALGYLS---DDSLFNKTKMPADIHLMAX 271
QY 458 -----EEMYPEYGVPPDPTQVE----- 475
DB 272 EIVRFHSIWPILMALDPLPKKVPFAHGMILMDGMSKSGKVVDVNNLIDRYGLDAT 331
QY 476 -----SGKGVF-----EGCYNYPDVVDLNNKNGKYGALBELYPLGNL 512
DB 332 RYVLMRELPGSDGVFTPEAFVERTNFDLAN-DLGNLVNRTISWINKYFDGEL 383

RESULT 13
US-10-234-026-13
;; Sequence 13, Application US/10234026
;; Publication No. US20030097679A1
;; GENERAL INFORMATION:
;; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
;; APPLICANT: Guy Jerome Corneel Bauw
;; APPLICANT: Mark William Davey
;; APPLICANT: Jens Ostergaard
;; APPLICANT: Marc Charles Ernest Van Montagu
;; TITLE OF INVENTION: PRODUCTION OF ASCORBIC ACID IN PLANTS
;; FILE REFERENCE: DECLES.001C1
;; CURRENT APPLICATION NUMBER: US/10/234,026
;; PRIOR APPLICATION NUMBER: NL 1006000
;; PRIOR FILING DATE: 1997-05-07
;; PRIOR APPLICATION NUMBER: PCT/EP98/02830
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 09/423,468
;; PRIOR FILING DATE: 2000-02-15
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 13
;; LENGTH: 600
;; TYPE: PRT
;; ORGANISM: *Braessica oleracea*
US-10-234-026-13

Query Match 3.1%; Score 94; DB 9; Length 600;
Best Local Similarity 17.9%; Pred. No. 6.9;
Matches 75; Conservative 55; Mismatches 128; Indels 160; Gaps 18;

QY 143 LGSHIVGGDGLIARLHGLPVMLSGVVYKPVLTEDSVLKVHNDSENGCELPWART 202
DB 210 IGGIIVGVAGHGTGARLP--PID---EQVIGMKLVTPAKGTIELSKD--NDPELFHAR 260
QY 203 GGGGNGFGIITKXYKFDLPMSPRGVASNLHFSWDFTRDALODLLTKYFKLARCWMKNT 262
DB 261 CGLGG--LSVVAEV---TLQCVBERQELLEHTYVS-----TLEIKGNHKLL----- 302
QY 263 VGNFOIFHOAAEFVWYLYTSYNDAREVAQODRHYLEADIEQIYKTCPTKALGCHAG 322
DB 303 -----STNKHVXYLYIPTYDV-----VVVTCNPVSKMSG--- 332
QY 323 WAFVPRPKRHTSKTSYVHDETMQDPFYALTEITNGSGPNQRKYSKAWMIKDPDFQI 382
DB 333 -----APKDKPKYTBELKHVRDLVR 354
QY 383 DVIMKYLTE-----VDPG-----LTSAMKDALLQVDMFG-----GEIHKVWMDATAVA 426
DB 355 ESTVKTQVODSSKKTTPDSREPDINELSFTELRDKLIADPLNDVHGVKNQ-----A 406
QY 427 QREYIIKLO-YQTYWQEE--DKDAVNLKWIIRDYEEMYPEYGVPPDPTQVESGKGVFEG 483
DB 407 EAEFMKSSSEYRAGWDELIGFDGCGQGVV---SETCFPAGTILAKPSMK----- 452
QY 484 CYNYPDVVDLNNKNGKYGALBELYPLGNLNLKAKMLDNPNI---FTNKOSIPTKP 538
DB 453 -----DLEYIEQLKELIQKALIPAPSPTEQRTGRSKSPWSP 489

RESULT 14
US-09-815-242-13013
;; Sequence 13013, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; PRIOR APPLICATION NUMBER: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 1410
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 13013
;; LENGTH: 657
;; TYPE: PRT
;; ORGANISM: *Staphylococcus aureus*
US-09-815-242-13013

Query Match 3.1%; Score 94; DB 10; Length 657;
Best Local Similarity 19.1%; Pred. No. 79; Mismatches 139; Indels 144; Gaps 21;
Matches 79; Conservative 51; Mismatches 139; Indels 144; Gaps 21;

```
QY 209 FGIITIKYKID.PMSRQVIAASNLHFSWDGFTFRLADLITKFKLARCQDKNTVG---- 264
DB 6 FFIITPIYV-----PSG-----NLHIGHAVST--VAGDIYARKRMQGYDVRLTGTDBH 53
QY 265 --KFOIFHOAA-----EEFY-----WYLYTSYNDABREVAQORHHLBADIYQIK 309
DB 54 GOKIQKAKKAGKTEIEYIDEMAGIKQIMAKLEISNDQFIRTEERHNV--VEQVBE 110
QY 310 TC--BETKALGSHAGWAPPPVRPRKHTKTSYMHDETMVPPYALTEITINGSQPORQK 367
DB 111 RLKQGDYILGSEYEGVSP-----DETYYTESQLVD--PQYENGKILIGKSPDSGHE 161
QY 368 YKSAYMIKDFPDF-----QIDYIMKYLTVPDGLTSAEMKDALLO-----VDM----- 410
DB 162 VE---LVKESYFNPISKYTDRLFEYDQNPDIQPPSRKNEMINNFIRKGLADLAVSRT 218
QY 411 -FGGEH-----KVVMDATAVAQREYIIKLOQYOTYQOEEDKQAVNLKWRDFF---- 457
DB 219 SFNWGVHVPSPKHHVVYVIMDALV--NYISALGYLS---DDESLFNKYPADIDHLMK 271
QY 458 -----EEMTEPYGVGVPDPTOVES----- 475
DB 272 EIVRHSIITPILMALDPLPKKVPFAHGMILMKQSKSKGNVDPVILIDRYGLDAT 331
QY 476 -----SGKGVF-----EGCYFNPYDVLNNMKXGKYGALFLYFLGNL 512
DB 332 RYIMRELPGSGDGVTFPEAFVERTNFDLAN-DLGNLVRTISMVWKYFDGEL 383
```

RESULT 15
US-09-815-242-5013
Sequence 5013, Application US/09815242
Patent No. US20020064569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5013
LENGTH: 584
TYPE: PRT
ORGANISM: Enterococcus faecalis

US-09-815-242-5013

Query Match 3.1%; Score 92; DB 10; Length 584;
Best Local Similarity 21.3%; Pred. No. 10; Mismatches 118; Indels 128; Gaps 18;
Matches 77; Conservative 38; Mismatches 118; Indels 128; Gaps 18;

```
QY 216 YEKDLPMSRQVIAASNLHFSWDGFTFRLADLITKFKLARCQDKNTVGKFOIFHOAAE 275
DB 191 HFYALPQSPQ--LFRQL-LMGAGFDR-----YQIVRC----- 220
QY 276 FMYLYTSYNDABREVAQORHHLBADIYQIKT-----CEPTKALGSHAGW 323
DB 221 ---FRDEDLRGDRQBEFTQ-----IDLETFLPPEEIQYTEMNMLAEVMKFTKIEIS 270
QY 324 APP-----VPRKHTKTSYMHDETMVPPYAL-----TETINGSQ 361
DB 271 VEPFPMSTDEAMARKSGDKPDRFAMELIDVADVAKDVDFKVFQALENGSHVKAALNAG 330
QY 362 PNQR-----GKYSAYMIKDFPDFOID-----VIMKYLTVPDGL--TSNEMKO 403
DB 331 AADKYSRKMDMLGKVVSGFGAKGLAMLKVEEDGKPIAKFLTEVSDELIAATNAVEGD 390
QY 404 ALLQVDMFGGEIHKVYWDATAVAQREYIIKLOQYOTYQOE--DKDAVNLKWRDFF-- 456
DB 391 IL---MFGADKPEIYVAAALGAV-----RTRLGKELGLIDESKENFLMIYDWPFL 436
QY 457 -YEEMTEPYGVGVPDPTOVES-----GKGFEGCYFNPYDVLNNMKXGKYGALFL 506
DB 437 EYDEEAGRYVSAHHPTQPKADVARLATDPASVIAEAY---DVLNNGYEJGG--GSLKI 491
QY 507 Y 507
DB 492 H 492
```

Search completed: June 10, 2003, 10:07:16
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:58:02 ; Search time 18 Seconds
(Without alignments)
892.495 Million cell updates/sec

Title: US-09-998-284-2

Perfect score: 2997
Sequence: 1 MATLPQKDPGVIVDVNAGT.....IFTKQSIPTKPKPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5B COMB.pep.*
3: /cgn2_6/prodata/1/1aa/6A COMB.pep.*
4: /cgn2_6/prodata/1/1aa/6B COMB.pep.*
5: /cgn2_6/prodata/1/1aa/PCUS COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2989	99.7	546	4	US-08-669-304-31
2	635.5	21.2	539	4	US-09-036-987A-11
3	635.5	21.2	539	4	US-09-370-700-11
4	323	10.8	397	4	US-09-433-248A-2
5	304	10.1	529	4	US-09-433-248A-6
6	275	9.2	538	4	US-09-647-390-16
7	266	8.9	540	4	US-09-647-390-18
8	255	8.5	495	4	US-09-217-490-2
9	117	3.9	129	4	US-09-433-248A-8
10	107	3.6	19	4	US-08-676-186-1
11	95	3.2	622	4	US-09-311-626B-4
12	94	3.1	600	4	US-09-423-468A-13
13	92	3.1	466	3	US-08-704-711A-17
14	92	3.1	466	4	US-09-521-220-17
15	92	3.1	944	4	US-08-652-877-12
16	92	3.1	4655	4	US-08-652-877-84
17	91.5	3.1	858	4	US-09-255-829-22
18	91.5	3.1	858	4	US-09-255-829-29
19	91.5	3.1	1169	4	US-09-255-829-20
20	90.5	3.0	1861	2	US-08-790-912-4
21	90.5	3.0	603	4	US-09-134-001C-5226
22	90.5	3.0	715	4	US-09-620-412C-329
23	90	3.0	23	4	US-08-669-304-2
24	90	3.0	23	4	US-08-669-304-9
25	89.5	3.0	4655	4	US-08-652-877-90
26	89.5	3.0	444	1	US-09-178-002-2
27	89.5	3.0	467	1	US-09-178-002-4

28	89.5	3.0	467	4	US-09-391-104-24	Sequence 24, Appl
29	89.5	3.0	468	4	US-08-448-489-13	Sequence 13, Appl
30	89.5	3.0	511	1	US-08-480-604A-20	Sequence 20, Appl
31	89.5	3.0	511	2	US-08-405-496A-20	Sequence 20, Appl
32	89.5	3.0	511	4	US-08-915-136-20	Sequence 20, Appl
33	89.5	3.0	511	4	US-08-957-310-20	Sequence 20, Appl
34	89.5	3.0	608	1	US-08-480-604A-21	Sequence 21, Appl
35	89.5	3.0	608	2	US-08-405-496A-21	Sequence 21, Appl
36	89.5	3.0	608	4	US-08-915-136-21	Sequence 21, Appl
37	89.5	3.0	608	4	US-08-957-310-21	Sequence 21, Appl
38	89.5	3.0	609	1	US-08-480-604A-30	Sequence 30, Appl
39	89.5	3.0	609	4	US-08-915-136-30	Sequence 30, Appl
40	89.5	3.0	2366	1	US-08-480-604A-10	Sequence 10, Appl
41	89.5	3.0	2366	2	US-08-405-496A-10	Sequence 10, Appl
42	89.5	3.0	2366	4	US-08-915-136-10	Sequence 10, Appl
43	89.5	3.0	2366	4	US-08-957-310-10	Sequence 10, Appl
44	89	3.0	751	4	US-08-969-415-2	Sequence 2, Appl
45	88	2.9	383	4	US-09-434-774-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-669-304-31
Sequence 31, Application US/08669304
Patent No. 6251626
GENERAL INFORMATION:
APPLICANT: Peter Stougaard
APPLICANT: Ole Cal Hansen
TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hunton & Williams
STREET: 1900 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006-1109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/669,304
FILING DATE: 12 July 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,910
FILING DATE: 7 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Stanislaus Arstman
REGISTRATION NUMBER: 28,562
REFERENCE/DOCKET NUMBER: 320,000003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 955-1926
TELEFAX: (202) 778-2201
TELEX: No. 6251626e
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-669-304-31
Query Match 99.7%; Score 2989; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 3.1e-296;
Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 MATLPQKDPGVIVDVNAGTDPKRLPSMKQGFNRWIGTNIIDFVVVYTPQACTAL 60

```

Db      1 MATLPQDQPGYIVDVNAGTADKDPRLPSMKQGFNRMWIGTNDIFYVYVYTPQAGTAL 60
Qy      61 DRAMEKSPGTIVRVSGCHCYEDFVPEBCYKAIINVTGLVSGYDDDRGIFVSSGDTNMG 120
Db      61 DRAMEKSPGTIVRVSGCHCYEDFVPEBCYKAIINVTGLVSGYDDDRGIFVSSGDTNMG 120
Qy      121 SFKTLFRDGRVLPVSGSCYSVGLGHIYVGGDGLIARLHGLPYDMLSGVEVYVVKPVLTE 180
Db      121 SFKTLFRDGRVLPVSGSCYSVGLGHIYVGGDGLIARLHGLPYDMLSGVEVYVVKPVLTE 180
Qy      181 SVLKYYKDSGNDGELFMAHTGGGGNGFGIITKYYFKDLPMSPRGVIASNLHFSMDGFT 240
Db      181 SVLKYYKDSGNDGELFMAHTGGGGNGFGIITKYYFKDLPMSPRGVIASNLHFSMDGFT 240
Qy      241 RDALQDLITKFKLARCDCMKTVGKFOIFHOAAEFWMLYTSNSDAEREVADRRHYL 300
Db      241 RDALQDLITKFKLARCDCMKTVGKFOIFHOAAEFWMLYTSNSDAEREVADRRHYL 300
Qy      301 EADILOIYKTCPTKALGSHAGMAPFVPRPKRHTSKTSYMHDETMDYPPFALTETINGS 360
Db      301 EADILOIYKTCPTKALGSHAGMAPFVPRPKRHTSKTSYMHDETMDYPPFALTETINGS 360
Qy      361 GPNRGKYSAYMIKDPDFQIDVIWXYLLEVPDGLTSAEMKDALQVDMFGGEIHKVW 420
Db      361 GPNRGKYSAYMIKDPDFQIDVIWXYLLEVPDGLTSAEMKDALQVDMFGGEIHKVW 420
Qy      421 DATAOREYIKIQTOTYMOEEDKAVNLKIMDFEEMTEPFGVDPDPTQVESGKV 480
Db      421 DATAOREYIKIQTOTYMOEEDKAVNLKIMDFEEMTEPFGVDPDPTQVESGKV 480
Qy      481 FEGGYFNPVDLNNMKXGKGALELYFLGNLRLIKAKMLMDNEIFTNKOSIPTKPLK 540
Db      481 FEGGYFNPVDLNNMKXGKGALELYFLGNLRLIKAKMLMDNEIFTNKOSIPTKPLK 540
Qy      541 EPKQTK 546
Db      541 EPKQTK 546

RESULT 2
US-09-36-987A-11
; Sequence 11, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-36-987A-11

Query Match      21.2%; Score 635.5; DB 4; Length 539;
Best Local Similarity 30.2%; Pred. No. 5e-56;
Matches 174; Conservative 91; Mismatches 216; Indels 95; Gaps 16;

Qy      2 ATLPQDQPGYIVDVNAGTADKDPRLPSMKQGFNRMWIGTNDIFYVYVYTPQAGTALD 61
Db      25 APNNRRITPTEI-----TVEPDDPEYPLVGNHNRFTG-KERIHIASADVVHAVA 77
Qy      62 RAMEKSPGTIVRVSGCHCYEDFVPEBCYKAIINVTGLVSGYDDDRGIF-VSSGDTNMG 120
Db      78 DAVR--TGRVVGSRSGCHGFENLVADPAIRVLVDLSLRVYDSTRGAFALBAGALGQ 135
Qy      121 SFKTLFRDGRVLPVSGSCYSVGLGHIYVGGDGLIARLHGLPYDMLSGVEVYVVKPVLTE 180
Db      136 VYRLEFRKMGVTTPTACPGVAGGHIIGGGYGLSRFRGSSVVDYLGVEVYVVDQAGEV 195
Qy      181 SVLKYYKDSGNDGELFMAHTGGGGNGFGIITKYYFKD-----LPMSPRGVIA 229
Db      196 HIVE-ADRNSTGAGHLMWMAHTGGGGNGFGIYTRFLRPDVVSTDAABELPRPATVLL 254
Qy      230 SNLHFSMDGFTRDALDLTKYFKLARCDCMKTVGKFOIFHOAAEFWMLYTSY----- 284
Db      255 RSFHWPMHLETSQSPALLQNF-----GNWYEOHSAPESTOGLFSTLVCAHR 302
Qy      285 -----SNDAREEVADRRHYLEADIEOYKTCPTKALGSHAGMAPFVRP 330
Db      303 QAGYVTLNHLDTDPNAERTLAE---HLSAINAOVGT--PAEGLRETLPL----- 350
Qy      331 RKRHTSKTSYMHDETMDYPPFALTETINGS GPNRGKYSAYMIKDPDFQIDVIWXYLT 390
Db      351 --RSTQVAG-----ATAE--GGEPPGQRTKVAAYLRGTGSEALTVVRRLT 394
Qy      391 EVPDGLTSAEMKDALQVDMFGGEIHKVWDATAVAQREYIKIQTOTYMOEEDKAVNL 450
Db      395 ----VGYDNPAAALLLLGCGMANAVAPSAATALQORSVLKALEFVTWNSPEADERHL 449
Qy      451 KWIIRDFEEMTEPFGVDPDPTQVESGKVFEGECYFNYPVDLNN--WKXGKGALELYF 508
Db      450 TWIRGFREMYAFTGGVPVPGTRV-----DSYINYPDTLADPLMNTSGVAMHLY 502
Qy      509 LGNINRLIKAKMLMDNEIFTNKOSIPTKLPKPKQ 544
Db      503 KDNVPLQRAKARMDPQNIQHGSLIKPARLSPGQ 538

RESULT 3
US-09-370-700-11
; Sequence 11, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1

```



```

Db 65 NPRWNISRRKPLMLTPFHESEIOAAIL-----CSKEIKLOLRVRSRGHDYEGLSY 116
Qy 87 DECVCAL-----INVTGVESGYDDRRGYF---VSSGDTNMGSPFKTLFRDGRVLPGSSCY 139
Db 117 LSDPFPVWDLINRS-TEINLADETAMVQAGSIGELYKISGA-SKVHG--PPACTCP 172
Qy 140 SVGLGHIIVGGGDGILLARLHGLPVDMLSGVEVVVKPVLTEDSVLKVVHKSSENGDELFW 199
Db 173 SVGIGHIISGGGGGLMKRKGILADNVDAVLI-----DANGKIHDRSMGED--VFW 223
Qy 200 AHNGGGGANGIITKYFYK-----DLNPSPGVIANLHFSWGFTRDALODL 248
Db 224 AIRGGDASSFGVILAMKIKLVPRPIVTFGNVPRTPPEEGVTDLIH-FMOYIAHDLHEDLV 282
Qy 249 TKYFKLARCMKXNVGKQ-----IFHOAAEFMYLYTSYNDABEVAQDRHYHEAD 303
Db 283 IRV-IAQISGHDSSKFFRATENSIFLGVDRLIPLNSESF-----ELGLQA- 328
Qy 304 IEQIYKTC-----EPTKALGHAAGWAPFVPRPKRHTSKTSYMHDETMDFPYALTETI 357
Db 329 ----KCTEMSWIQSVMFIAGYNIEDPLLELLNRTTMFKRSF----- 366
Qy 358 NGSGPNRGKYYKAYMIKDPDPFOIDVIWXYLTVEPDGLTSAEMKDALLOVDFEGEIHK 417
Db 367 ----KAKSDPFKEBPVPSGLLEGAMKIL-----LEBEIAFLIMEPYGGRME 408
Qy 418 VVMDATAVAGRE-YIILQYQTVWQEDKAV--NLKMRDPEEMVEPYGVDPPTQV 474
Db 409 ISESEIPFPHKGNLYMLQYLVNNEVSDEASRHLQWAKVAYKM-TPV-----V 458
Qy 475 ESGKGVFEGCYFNPVDVL--NNMKNKYGALDEL---YFLGNINRLIKAKMLDPNEIF 528
Db 459 SKSP---RAAYFNKCDLDLGNKJLDTSYSSEASVWGKKYFKGNRRRLAQIKTKFDPINF 515
Qy 529 TNKQSIIP 535
Db 516 RNEQSIP 522

```

RESULT 6
US-09-647-390-16
; Sequence 16, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
; APPLICANT: Stuijver, Maarten
; APPLICANT: Simons, Lambertus
; TITLE OF INVENTION: Pathogen-Inducible Promoter
; FILE REFERENCE: MOG 57707/UST
; CURRENT APPLICATION NUMBER: US/09/647,390
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: EP 98201024.1
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: PCT/EP99/02178
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Helianthus annuus
US-09-647-390-16

Query Match 9.2%; Score 275; DB 4; Length 538;
Best Local Similarity 23.9%; Pred. No. 3,1e-10;
Matches 144; Conservative 64; Mismatches 197; Indels 198; Gaps 26;

```

Qy 22 DKPPD-----RLPSMKOGF--NRRWIGNIDPVVYVYVPOG-----AC 57
Db 43 DRADPSFPIGEVYTPGNSFFVYLQNTYIRNLKRNENETTPRPFLITIAEHVSHIOAAVVC 102
Qy 58 TALDRAMKCSPTGTVRIVSGHCYED-----FVPECVKALINV-----TGLIVE 101

```

```

Db 103 GKQNRLLKTR-----SGCHDYEGLSYLTNTNPPFIVDMFNLSINVDEOETAMVQ 155
Qy 102 SGYDDDRGYFVSSGDTNMGSPFKTLFRDGRVLPGGSYVSGVGHGVGGDGLLARLHGL 161
Db 156 AGATLGEVYIRIAEKN-----KHG--FPAGVCPVVGCHHSGGSGYGNLARKYGL 204
Qy 162 PVDMLSGVEVVVKPVLTEDSVLKVVHKSSENGDELFWAHTGGGGNGFIITKYFKDLP 221
Db 205 SVDNIYDAQI-----DVNGKILDRKSMGED--LFWAITGGGVSFGVLAAYKIK-LV 254
Qy 222 MSPRGVIANLHFSWGFTRDALQDLITKYFKLARCDMKQYTKFOIFHOAAEFMYLY 281
Db 255 RVEEVTV-----FTIERREON----- 272
Qy 282 TSYSDNAER--EVAQ--DRHYHEADIEQIYKTCPTKALGHAAGWAPFVPRPKRHTSK 337
Db 273 --LSTIAERWVQVADLDBDLFLRMTPSVINDT-----NGKTVAAIIP-- 314
Qy 338 TSYMHDE-----TMDYFPYALTE-----TINGSGPNOR- 365
Db 315 TLVLGNSRNLVTLNKPPELGLQESDCTEMSWVESVLYTGPPSGPTTALLSRPRL 374
Qy 366 --GKYSAYMIKDPDPFOIDVIWXYLTVEPDGLTSAEMKDALLOVDFEGEIHKVM DAT 423
Db 375 NPKIKSDYQNPDISKROEFIFERL-----KELNOMLAFNPYGRMSISEFAK 425
Qy 424 AVAGRE-YIILQYQTVWQEDKAV--LKMIRDPEEMVEPYGVDPPTQVESGKV 480
Db 426 PPHRSGNIAKIQYENWEDLSDEAENRYLNFRLMYDM-----TPYSKMRK-- 475
Qy 481 FEGCYFNPVDVLNNMKNKYGALDELVPLG-----NLNRLIKAKMLDPNEIFTNKOS 533
Db 476 ---AFINYDDLDIGINSHGRNAVTEGMVYGHKYFKETNYKRLVSVTKVDPDWFNBOG 532
Qy 534 IPT 536
Db 533 IPT 535

```

RESULT 7
US-09-647-390-18
; Sequence 18, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
; APPLICANT: Stuijver, Maarten
; APPLICANT: Simons, Lambertus
; TITLE OF INVENTION: Pathogen-Inducible Promoter
; FILE REFERENCE: MOG 57707/UST
; CURRENT APPLICATION NUMBER: US/09/647,390
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: EP 98201024.1
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: PCT/EP99/02178
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Lactuca sativa
US-09-647-390-18

Query Match 8.9%; Score 266; DB 4; Length 540;
Best Local Similarity 23.5%; Pred. No. 2,6e-18;
Matches 140; Conservative 72; Mismatches 187; Indels 198; Gaps 30;

```

Qy 20 TPKPPRLPSMKOGFNRRWIGNIDPVVYVYVPOG--ACTALD---RAMEKCSPTV 73
Db 56 TPD--NSSPFSVQAYIR-----NLRFNEST--TPKILITIALHPHIOAAVVCAR--THR 106
Qy 74 IV-----SGHCYED-----FVPECVKALINVGLVESGYDDRRGYFVSSGDTN 118

```


RESULT 10
US-08-676-186-1
Sequence 1, Application US/08676186
Patent No. 6358543
GENERAL INFORMATION:
APPLICANT: SOE, John B.
POULSEN, Charlotte H.
HOSTRUP, Pernille B.
TITLE OF INVENTION: A method of improving the
properties of a flour dough, a flour dough improving
composition and improved food products
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 1850 K Street N.W., Suite 450
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-2296
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,186
FILING DATE: 12-Sep-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/483,870; PCT/DK96/00239
FILING DATE: 07-JUN-95; 04-JUN-96
ATTORNEY/AGENT INFORMATION:
NAME: Aksman, Stanislaus
REGISTRATION NUMBER: 28,562
REFERENCE/DOCKET NUMBER: 44648/013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-778-8300
TELEFAX: 202-778-8335
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-676-186-1
Query Match 3.6%; Score 107; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 DPGYIVDVNAGTPDKPDP 26
Db 1 DPGYIVDVNAGTPDKPDP 19

RESULT 11
US-09-311-626B-4
Sequence 4, Application US/09311626B
Patent No. 6399347
GENERAL INFORMATION:
APPLICANT: Jorgensen, Per Lina
APPLICANT: Schmor, Kirk
APPLICANT: Andersen, Lene No. 6399347boe
APPLICANT: Schuelein, Martin
APPLICANT: Outtrup, Helle
TITLE OF INVENTION: No. 6399347el Rhamnogalacturonan Hydrolases
FILE REFERENCE: 5572,204-US
CURRENT APPLICATION NUMBER: US/09/311,626B

CURRENT FILING DATE: 1999-05-13
PRIOR APPLICATION NUMBER: 0608/98
PRIOR FILING DATE: 1998-05-01
PRIOR APPLICATION NUMBER: 60/084,358
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 622
TYPE: PRT
ORGANISM: Bacillus licheniformis
US-09-311-626B-4
Query Match 3.2%; Score 95; DB 4; Length 622;
Best Local Similarity 20.4%; Pred. No. 0.94;
Matches 114; Conservative 81; Mismatches 212; Indels 152; Gaps 33;
QY 31 MKQGFNRN---IGNIDFVYV---YTPQ---GACTYLDRAKCSGTIVRIVSGHC 80
Db 1 MKKG-KRWKNLLASLILITLVGFSEQAEADGRTAAQAROMESLNLGLVAVKTGNGV 59
QY 81 YEDF--VPDECVKATINV---TGLVESGYDDRGYFVSSGDTNMGSFITLFRDHGRVLP 134
Db 60 FVSWRLGTEPSSVSLNVRNGKLLGSPITSTNYQAGDLN----- 103
QY 135 GGSQSVGLGHIIVGGD-----GILARL-GLPVDMLSGVEVYVKPVLTEDSVLKTYV 186
Db 104 --AVYVRA---VLNREQAPSESIVGLNKQKSVPLQKPAQK-----TPDGV-SYT 150
QY 187 HDSQSGNDELFWAHTGGGCGNFITTKYFKDLPMSPRGVLASNLHFSWDGFTRLALOD 246
Db 151 YSANDASLDGLV---GDAQYEIFLKW-----DPSNSKONSQDGYTGDLVLD 193
QY 247 LITKYFKL-ARCOMKNTVKKFQIFHOAAEFVWYLYTSYNSDAEREAQDRHYHLADIE 305
Db 194 A---YKLDGTMWRINLGR---NIRGAHYTOFLVYDGDGKAEIAM----- 235
QY 306 QIYKCEPTK-ALGHAGWAPFVR--PRKRHTSKTSYM-----HDETMDYP----- 349
Db 236 --KTADGRTKDGKVIYGNANDYRNAQRIISGPEYLVFKPDGTGAEILTVVYEPARGN 292
QY 350 -----FALITETINGSQPN---QRGKKSAMIK-DPPDQIIVIMKYLIE 391
Db 293 VADWGDYGNRVDRFLAGVAYLDERPSPFMAGYTRTVLVAVYNGGGLTLTMTFDS 352
QY 392 VP-DGLTSAEMKDALLOVMPFGEHKKVWDATAVAQREYITLQVQTYQOEED-DAVAV 449
Db 353 APGNAGYAGGNSHLSVADVDDGDKDEITYGMAV---DHDGKGLYSTGNGHDAHTGN 409
QY 450 LKWIR-----DFYEMYPEYG-GVDPDNT-----QVESGKGVFEGCYFNYPDVT----- 492
Db 410 LDSPRQGLEVFQYHENSNSPYGLSPFDATGKIWGVHAKDVGRCM---AADIDPRYEG 466
QY 493 LNNWKGKYGALDELFLGN 511
Db 467 AEVWANGSLYTAQVXIGN 485

RESULT 12
US-09-423-468A-13
Sequence 13, Application US/09423468A
Patent No. 6469149
GENERAL INFORMATION:
APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
APPLICANT: Guy Jerome Corneel Bauw
APPLICANT: Mark William Davey
APPLICANT: Jens Ostergaard
APPLICANT: Marc Charles Ernest Van Montagu
TITLE OF INVENTION: PRODUCTION OF ASCORBIC ACID IN PLANTS
FILE REFERENCE: DECLES,001APC
CURRENT APPLICATION NUMBER: US/09/423,468A
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: NL 1006000

```

; PRIOR FILING DATE: 1997-05-07
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Brassica oleracea
US-09-423-468A-13

```

Query Match	3.1%	Score 94	DB 4	Length 600	.
Best Local Similarity	17.9%	Pred. No. 1.1			
Matches 75, Conservative	55	Mismatches 128	Indels 160	Gaps 18	

Qy	143	LGSHIYGGSDGILARLHGPRVMDLSGVEVVPVLTETDSVLKYKHSBGNDGLFMAHT	202
Db	210	IGSHIYGVANGHGARGLR--PID-----EYIGMKLVTPRAKGIITELSKD---NDPELFILAR	262
Qy	203	GGGGNGFGIITKYKFDRLPMSPRGVIAINLHFSMDGFTRDALODLLTKYFKLARCDWKNT	262
Db	261	CGIGG--LGVAEY--TLQCVREJELLEHTYVS-----TLEIKKNHKKLL-----	302
Qy	263	VGRFOIYHQAEEFPWYLYLTSYSNDAERVAQDRHYLEADIYQIKTCETKALSGHAG	322
Db	303	-----STNKHVYKLYIPTYDTV-----VVVTCNVSRKMSG---332	
Qy	323	MAEPVPRPKRHTSKTSYMHDETMVYFYALTETINGSQPNORGKYSAYMIKDPPEQI	382
Db	333	-----APDKREKYTEELAKVRLYLIR	354
Qy	383	DVIWKYLT-----VDDG-----LTSAMKDALLOVDMFG-----GEIHKRVMDATAVA	426
Db	355	ESIIVKRVYDSSKKRPPDSRREPINELSPFELDKLIAIDPLNDVHYGVKNQ-----A	406
Qy	427	QREYIIKIQ-YQTYMOE--DKDANVLMKIRDPFYEMTEEPYGGVDPRTTVQVESGKVPEG	488
Db	407	EAEFMKKSGBYRGVMSDETLGPDCCGQOQV-----SETCEPAGTLAKPSMK-----	452
Qy	484	CYENYVDVDLNMWKGKYGALLETFLYGLNLRILAKMLMDNEI-----FTNKGSIPTKP	538
Db	453	-----DLEIETEOUKELIQCEALPASPPIORFTGSKSGKSPMP	489

RESULT 13
 US-08-704-711A-17
 ; Sequence 17, Application US/08704711A
 ; Patent No. 6114159
 ; GENERAL INFORMATION:
 ; APPLICANT: WILLY, Horst
 ; APPLICANT: HINZMANN, Bernd
 ; TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX
 ; TITLE OF INVENTION: METALLOPROTEASES, THEIR PRODUCTION AND USE
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; City: Washington
 ; STATE: D. C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/704,711A
 ; FILING DATE: 20-NOV-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/DE95/00357
 ; FILING DATE: 17-MAR-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: DE 4438838.1

? FILING DATE: 21-OCT-1994
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: DE 4409663.1
 ? FILING DATE: 17-MAR-1994
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: GRANAPOS, Patricia D.
 ? REGISTRATION NUMBER: 33,683
 ? REFERENCE/DOCKET NUMBER: 26083/124
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (202)672-5300
 ? TELEFAX: (202)672-5399
 ?
 ? TELEX: 904136
 ? INFORMATION FOR SEQ ID NO: 17:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 466 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ?
 ? US-08-704-711A-17

Query Match	3.1%;	Score 92;	DB 3;	Length 466;
Best Local Similarity	20.2%;	Pred. No. 1.2;		
Matches 100; Conservative	52;	Mismatches 158;	Indels 184;	Gaps 27;

Qy	94	INVTGLVES-----GYDDRGVFSGGDTNMSSFKTLFPDHDGRLVGGSCYCVG	142
Db	72	LNVTGKPNBEETLDMKKKPRCGVPRDSGGMILGPNKMKERTNLTYIRN-----	119
Qy	143	LGSHIVGGGDIILARLHGLPVDWLSGVEVVKPVLTEDSVLKYVHKDSEGNDELFWA--	200
Db	120	-----YTPOLSEAEVERAL-----KDAFELMSVA	143
Qy	201	---HTGGGGGNFGIITKYFKD---LPM-SPRGVIANSHFSWDGFTDALQDILLTX	251
Db	144	SPLIFTRISQEAENINIAFYORDHGDNSPFGPNGLILHAFO-PEQIGIGDAHF-----	197
Qy	252	FKIARCMKNTVAKFQIFHQAEEFVMVLNYSYSD-----AERVAODRHYLEA	302
Db	198	---AEEETNTNSANYNIFLVAANEHSGSLGLAHSSDPGALMYPNAFRETSS---NYSLQ	251
Qy	303	D-----IEQYKTCSEPTALGGHAGAPRPVRPKKHTSKTSMHDETMQYFVALTETLN	358
Db	252	DDIDIGIQLIY-----GLSSNPLOTPGTPPKPC--DPSL--TDAIT-----	289
Qy	359	GSGENORGK--YKSAYWIKDPDFQO-----IDVIKYULEVPRDGLTSA-EMKALLQV	408
Db	290	----TLNGEILFFDQRYFWRNRHNPQLQRVENMFLISLFWLSL--PLGIGQAYDHPBDL-I	341
Qy	409	DMFGEIHKV--MD-----ATAVAQREYTIKLQYOTYWOEBKDAVNLMK	452
Db	342	FLFGKNQYMAISGYDILQGYPKDISNYSYGFSSVQALDAVAFYSKTYFFVNDQ-----PW	396
Qy	453	IRDYEEHYER-----YGVPRDPTQVYESGKGVREGCYF-----NYPDVL-----	493
Db	397	RYDQORFMEGYPKSGISGAFPG---IESDAVAFQOEHFFHVFGSPRYAADFLLAQRYT	452
Qy	494	-----NNMKNGKYG	502
Db	453	RVANGNKMLNCRYG	466

RESULT 14
US-09-521-220-17
Sequence 17, Application US/09521220
Patent No. 6399348
GENERAL INFORMATION:
APPLICANT: WILL, Horst
HINZMANN, Bernd
TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX
METALLOPROTEINASES, THEIR PRODUCTION AND USE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/521,220
FILING DATE: 08-Mar-2000
CLASSIFICATION: <Unknown>
21-OCT-1994
17-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/704,711
FILING DATE: <Unknown>
APPLICATION NUMBER: DE 4438838.1
FILING DATE: 21-OCT-1994
APPLICATION NUMBER: DE 4409663.1
FILING DATE: 17-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: GRANADOS, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 26083/124
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 466 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-521-220-17
Query Match 3.1%; Score 92; DB 4; Length 466;
Best Local Similarity 20.2%; Pred. No. 1.2;
Matches 100; Conservative 52; Mismatches 158; Indels 184; Gaps 27;
QY 94 INVTGLVES-----GYDDRGVFSVSGDTNNGSFKTLPRDGRVLPGGSCYSVG 142
DB 72 LNVTKGKNEETLDMKKPRCGVPSGGFMTLPGRKERTVLTIRIN----- 119
QY 143 LGSHIVGGDGLARLHGLPVDMLSGVEVVVVKVLTEDSVLKVYKDSGNDGELFWA-- 200
DB 120 -----YTPQLSEAEVERAL-----KDAPELMSVA 143
QY 201 -----HTGGGGGNFGIITKYFKD-----LPM-SFPGVIAASNLHFSWDGFTDALODLLTKY 251
DB 144 SPLFTFISQGEADINIAFYGRDHGDSPPDPGNGILAHAFQ-DGQIGGGAHPD----- 197
QY 252 EKLARCMKNTVKGFOIFHOAAEEFWVLYTSYND-----AREVVAQDRHYLSEA 302
DB 198 ---AETWTNTSANTYPLVAHAFHGSLSLAHSSDQALMPTVYAFRETS---NYSLPQ 251
QY 303 D-----IQIYKTCPTPALGSHAGWAPPVPRRRKHTSKTSYHDETMWYFFVALTETIN 358
DB 252 DDIDIGIQAIV-----GLSSNPDIQPTGSPSTPKPC---DPSL-TFPAIT----- 289
QY 359 GSGGNQKQ---YKSAWIKDFPDPQ-----IDVIMKYLTEVVDGLTSA-EMDALLQV 408
DB 290 ---TLGGEILFFKDRYFWRHRPOLQRYEMNFIPLFWPSTL---PTGIQAAYEDPRDL-I 341
QY 409 DMFGGEIHKVY--ND-----ATAVAQREYTIKLOQYQWQEEEDKXAVNLKW 452
DB 342 FLEFGNQWALSGLDILQGYPKDISNIGFPSSVQALDAVAYRSKTYFFVNDQ-----FW 396
QY 453 IRDFYEEMVYD-----YGGVDPDNTQVESGKGVFEGCYF-----NYPDVDL----- 493

DB 397 RYDNGQRFMEPGYPKSISGAFFG-----IESKDAVFOQEHFFVFGSGPRYYAFDLAQRVT 452
QY 494 -----NNMKNGKYG 502
DB 453 RVARGNKMVLCRYG 466
RESULT 15
US-08-652-877-12
Sequence 12, Application US/08652877
Patent No. 6187548
GENERAL INFORMATION:
APPLICANT: Aketstrom, Goran
APPLICANT: Juhlin, Claes
APPLICANT: Raek, Lars
APPLICANT: Crumley, Gregg R.
APPLICANT: Morse, Clarence C.
APPLICANT: Murray, Edward M.
TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
TITLE OF INVENTION: Thereof and DNA Encoding Same
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Rd., 3043
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426-0107
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: System 7.5.1
SOFTWARE: Word 6.0 (Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/652,877
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/15203
FILING DATE: 22-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,836
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,314
FILING DATE: 07-JUNE-1995
ATTORNEY/AGENT INFORMATION:
NAME: Savitzky, Martin
REGISTRATION NUMBER: 29,699
REFERENCE/DOCKET NUMBER: A1355E-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-454-3816
TELEFAX: 610-454-3808
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-652-877-12
Query Match 3.1%; Score 92; DB 4; Length 944;
Best Local Similarity 22.3%; Pred. No. 3.7;
Matches 53; Conservative 35; Mismatches 90; Indels 60; Gaps 12;
QY 325 PFPVPRKRTSK---TSYHDE---TMDYF-----YALTEINGSGNQGRKYS 370
DB 348 PDNVRIRKYNLSSERSBYQDEEYIAVDYDMDPDIGLSVYYTVRGEG-SRFAIKR 406
QY 371 AYMIKDPDF-----QIDVIMKYLTEVVDGLTS-----AEMDALLQVDMFG 412

Db 407 AYI---PNFESGRNLTVOEVDIKLYVMQ-PDGIADVGRHIYWSDVKNKRIEVAKLD 461
QY 413 GEIHKYVM-----DATAVAQREYIIKLOQTYWOEDKDAVNLKWIIRDfYEEMYPEYGG 466
Db 462 GRVRR--WLISTDLDQPAIAVNPKLGIMFTDNGKEPK--IESAMNG-----506
QY 467 VPDPTQVESGKGVEFGCYFNYPDPVDLNNWKNKGYGALBELYFLGNLNRLIKAKMLWP 524
Db 507 -EDRNILVFEDLGWPTGLSIDYLNNDRIYWSDFKEDVIEFTIKYDGTDRVIAKEANMP 563

Search completed: June 10, 2003, 10:00:51
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: June 10, 2003, 09:54:36 ; Search time 35 Seconds
(without alignments)
3214.335 Million cell updates/sec

Title: US-09-998-284-2

Sequence: 1 MATLPQKDGKGYIVDNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21: *
1: sp archaea: *
2: sp bacteria: *
3: sp fungi: *
4: sp human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2997	100.0	546	10 P93762	P93762 Chondrus cr
2	771.5	25.7	508	16 Q8Y3D0	Q8Y3D0 ralistonia s
3	736.5	24.6	685	16 Q8ZJ22	Q8ZJ22 yersinia pe
4	674.5	22.5	550	16 Q9EX55	Q9EX55 streptomyc
5	635.5	21.2	539	2 Q9ALN1	Q9ALN1 saccharopol
6	633.5	21.1	545	16 Q9J308	Q9J308 streptomyc
7	590.5	19.7	489	2 Q8VMA5	Q8VMA5 streptomyc
8	440.5	14.7	485	2 Q9ZAR8	Q9ZAR8 streptomyc
9	399	13.3	480	16 P71091	P71091 bacillus su
10	368.5	12.3	530	10 Q9FZC5	Q9FZC5 arabidopsis
11	368.5	12.3	530	10 Q949N1	Q949N1 arabidopsis
12	347.5	11.6	539	10 Q9SVG3	Q9SVG3 arabidopsis
13	346	11.5	528	10 Q9SVG7	Q9SVG7 arabidopsis
14	333.5	11.1	535	10 Q9AYM8	Q9AYM8 vigna ungu
15	328.5	11.0	529	10 Q9FZC8	Q9FZC8 arabidopsis
16	328.5	11.0	535	10 Q9FZC4	Q9FZC4 arabidopsis

17	324.5	10.8	541	10 Q9LPC3	Q9LPC3 arabidopsis
18	322	10.7	540	10 Q9SVG4	Q9SVG4 arabidopsis
19	317.5	10.6	534	10 Q9JZA3	Q9JZA3 arabidopsis
20	305	10.2	542	10 Q9FKU9	Q9FKU9 arabidopsis
21	304.5	10.2	552	10 Q8RWA4	Q8RWA4 arabidopsis
22	302.5	10.1	537	10 Q9FKU8	Q9FKU8 arabidopsis
23	302.5	10.1	552	10 Q9FZC7	Q9FZC7 arabidopsis
24	301	10.0	542	10 Q9JY11	Q9JY11 arabidopsis
25	300.5	10.0	532	10 Q64743	Q64743 arabidopsis
26	300	10.0	532	10 Q9SVG5	Q9SVG5 arabidopsis
27	298.5	10.0	527	10 Q9FZC6	Q9FZC6 arabidopsis
28	297	9.9	533	10 Q9F121	Q9F121 arabidopsis
29	296	9.9	536	10 Q9SA99	Q9SA99 arabidopsis
30	294.5	9.8	541	10 Q9FKV0	Q9FKV0 arabidopsis
31	278.5	9.3	447	16 Q06997	Q06997 bacillus su
32	278	9.3	533	10 Q9SA89	Q9SA89 arabidopsis
33	277.5	9.3	526	10 Q9SA88	Q9SA88 arabidopsis
34	277	9.2	527	10 Q9SA85	Q9SA85 arabidopsis
35	275.5	9.2	715	10 Q9LNL9	Q9LNL9 arabidopsis
36	274	9.1	532	10 Q9FKV2	Q9FKV2 arabidopsis
37	273	9.1	536	10 Q9ZPP5	Q9ZPP5 berberis et
38	273	9.1	535	10 Q9F125	Q9F125 arabidopsis
39	272.5	9.1	530	10 Q94KD7	Q94KD7 arabidopsis
40	272.5	9.1	530	10 Q9SUC6	Q9SUC6 arabidopsis
41	272	9.1	538	10 Q8SA59	Q8SA59 helianthus
42	266	8.9	540	10 Q8SA60	Q8SA60 lactuca sat
43	252	8.4	531	10 Q9SA86	Q9SA86 arabidopsis
44	251.5	8.4	464	2 Q9KHK2	Q9KHK2 streptomyc
45	249.5	8.3	540	10 Q64745	Q64745 arabidopsis

ALIGNMENTS

RESULT 1

P93762 PRELIMINARY; PRT; 546 AA.
ID P93762;
AC P93762;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hexose oxidase.
OS Chondrus crispus (Carrageen).
OC Eukaryota; Rhodophyta; Floridiophyceae; Gigartinales; Gigartinales;
OC Chondrus;
OX NCBI_TaxID=2769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97269074; PubMed=9111074;
RA Hansen O.C., Stougaard P.;
RT "Hexose oxidase from the red alga Chondrus crispus. Purification,
molecular cloning, and expression in Pichia pastoris."
RL J. Biol. Chem. 272:11581-11587(1997).
DR EMBL; U89770; AAB49376.1; -;
DR InterPro; IPR001575; Oxid_FAD_Bind.
DR Pfam; PF01565; FAD binding 4; 1.
SQ SEQUENCE 546 AA; 61899 MW; 8DDDEA46D53C870B1 CRC64;

Query Match 100.0%; Score 2997; DB 10; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.1e-222;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATLPQKDGKGYIVDNAGTPEKPRRLPSMKGFNRKMTGNTIDVVVYVYTPQACTAL 60
DB 1 MATLPQKDGKGYIVDNAGTPEKPRRLPSMKGFNRKMTGNTIDVVVYVYTPQACTAL 60
QY 61 DRAMEKSPGTVRIVSGGCHYEDFVPEDECVKAIINTVGLVSGYDDDRGFVSSGNTNM 120
DB 61 DRAMEKSPGTVRIVSGGCHYEDFVPEDECVKAIINTVGLVSGYDDDRGFVSSGNTNM 120
QY 121 SFTFLFRDHGRVLPGGSCYVGLGHIIVGGDGIARLHGLPYDWLISGVYVVKPYLTED 180
DB 121 SFTFLFRDHGRVLPGGSCYVGLGHIIVGGDGIARLHGLPYDWLISGVYVVKPYLTED 180

QY	181	SVLKVYHKS	SEGNDGELFWAHHTGGGGNGN	GIITKKYFPDLPMS	PRGVASNLHFSW	GFT	240					
Db	181	SVLKVYHKS	EGNDGELFWAHHTGGGGNGN	GIITKKYFPDLPMS	PRGVASNLHFSW	GFT	240					
QY	241	RDALODLLT	YFPLAEC	CDWKNTVGKFOI	FHOAAEFVWYLTSYS	ANDEREVAODRRYHL	300					
Db	241	RDALODLLT	YFPLAEC	CDWKNTVGKFOI	FHOAAEFVWYLTSYS	ANDEREVAODRRYHL	300					
QY	301	EADIEQIYKT	CEPTKALG	HAGWAPPVAPR	KRKHTSKTSY	NHDETM	DYPPAL	TEITNGS	360			
Db	301	EADIEQIYKT	CEPTKALG	HAGWAPPVAPR	KRKHTSKTSY	NHDETM	DYPPAL	TEITNGS	360			
QY	361	GPBORGYK	SAWIMKDP	PPFOIDVIM	KYLT	TEVPDGLT	SAEMKDALLOVDM	GGSEI	HRVW	420		
Db	361	GPBORGYK	SAWIMKDP	PPFOIDVIM	KYLT	TEVPDGLT	SAEMKDALLOVDM	GGSEI	HRVW	420		
QY	421	DATAVAQRE	YIILQYQ	YVMOEEDK	QAVNLK	IRDPYE	MEP	PGVADP	PTQV	ESGKV	480	
Db	421	DATAVAQRE	YIILQYQ	YVMOEEDK	QAVNLK	IRDPYE	MEP	PGVADP	PTQV	ESGKV	480	
QY	481	PEGCEYFN	YPPVDL	NNKNGK	YGALE	YFLGIN	NRLIKAK	WIMD	PNELFT	NKOSIPT	YPLK	540
Db	481	PEGCEYFN	YPPVDL	NNKNGK	YGALE	YFLGIN	NRLIKAK	WIMD	PNELFT	NKOSIPT	YPLK	540
QY	541	EPKQTK	546									
Db	541	EPKQTK	546									

RESULT 2

ID	Q8Y3D0	PRELIMINARY;	PRT;	508 AA.
AC	Q8Y3D0;			
DT	01-MAR-2002	(TREMBLrel. 20. Created)		
DT	01-MAR-2002	(TREMBLrel. 20. Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21. Last annotation update)		
DE	Hypothenelial protein Rsc0050.			
NR	RSC0050 OR RS01873.			
OS	Ralstonia solanacearum (Pseudomonas solanacearum).			
OC	Bacteria; Proteobacterium; beta subdivision; Ralstonia group			
CC	Ralstonia.			
NCBI	Taxid=305;			
XX				

RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cartolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaupin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigurter P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissensbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";
RL Nature 415:497-502(2002).
DR EMBL; AL646057; CAD13578.1; -.
DR InterPro; IPR001575; Oxid PAD bind.
DR Pfam; PF01565; PAD binding 4; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 508 AA; 56031 MW; ED8F025998BC3C84 CRC64;

Query Match	25.7%;	Score	771.5;	DB	16;	Length	508;
Best Local Similarity	35.2%;	Pred.	No. 4.2e-51;				
Matches 196;	Conservative	74;	Mismatches	202;	Indels	85;	Gaps
							17;

[illegible]

```

QY 129 HGRVLPGSSCSYVGLGGHIVGGSDGILRLRHGLPVDMLSGVEVVVXKPVLTEDSVLEKYVHK 188
Db 114 HGVTLPGSSCSYVAGGHHISGGGYGLLSRLOGLTVDMLSAVDIVT-----V 159
QY 189 DSEG-----NDGELFWAHTGGGGGNFGIITKYUYFKDLPMSPRGVIASNLHFSWDG 238
Db 160 DKQGRHAPRTVDADARPDLFRACRGAGGNGFIIITSYTFARJPEAPREVALATVAFDMAA 219
QY 229 FTFRDLQDLILKY--FKLARCDMKTGVKFOF---HQAAEFWVWLYLYSYNDAREVA 293
Db 220 MTPKEFALLRLYGDYMATRGKDPRTWGLFSLILKTHKSAGQIVM--LQPCNPD---G 273
QY 294 QDRNHYLEADIEQIYKTECEPTKALGSHGMAF-----PVPRKRHNTSKTSMHEDTMD 347
Db 274 TCRDLSVNLDFLARQVCAPVAPARAPRPGYGRANHQGVGLLSKPH-----VVO 324
QY 348 YFPVALTTTNGSGENORGKYSAMWIKDPFQIDVIMWKYLTVEYBDGLTSEMMDALLO 407
Db 325 YDMLTATOSLNGSGENORGKYSAMTKKGFTAREKORITHTLRTVQGI--DLQSLOL 381
QY 408 VDMFGGEIHKYVWDF--TAAOREYIITKLOYQTYWQEBDADANLKMIRDFEEMY----- 461
Db 382 VDSYGAIVAKBRLIDTAPORASIMKIOYQYTWISAADDAHLKIRIGFHYDVIQTPDV 441
QY 462 -EPYGVDPDPTQVESGKVEGECYFNPVDVDL---NNWKKGKYGALBELLYLGNLRLIK 517
Db 442 SAPHAGTEYDGR-----YEGCYINPDVDMILAVFPWQLYUGDGLYAF-----LQR 489
QY 518 AKMLMDNEIETNKOSI 534
Db 490 VKRRYDPNNITFHMAVSV 506

```

RESULT 3

ID	PRELIMINARY;	PRT;	685 AA.
AC	Q8ZJ22;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Putative oxidoreductase.		
GN	YP00308.		
OS	Yersinia pestis.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Yersinia.		
OX	NCBI_taxid=632;		

RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill U., Wren B.W., Thomson N.R., Titchell R.W., Holden M.T.G.,
RA Prentice M.B., Sebaldia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baaham D., Bentley S.D., Brooks K., Cedenio-Tarrega A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Letwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moutle S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,
RT "Genome sequence of *Yersinia pestis*, the causative agent of plague."
RL Nature 413:5233-5271(2001).
DR EMBL; AJ414142; CAC89170.1; "-
DR InterPro; IPR003610; CBM 5.12.
DR InterPro; IPR003961; PR T1T.
DR InterPro; IPR001575; OXLD_FAD_bind.
DR Pfam; PF02839; CBM 5.12; 1.
DR Pfam; PF01565; FAD_binding_4; 1.
DR Pfam; PF00041; fnj3 1.
DR SMART; SM0060; FN3; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 685 AA; 75266 MW; 597853869F857CC6 CRC64;

Query Match	24.6%	Score 736.5	DB 16	Length 685
Best Local Similarity	33.6%	Pred. No. 3.2e-48		
Matches 195; Conservative	77	Mismatches 205	Indels 103	Gaps 20

QY 364 QGKGT--KSAVMIKDFPDFOIDVIMKYLTVEPDLGTSAMKDALLOVDNFGGEIHKRVWD 421
DB 380 GGRFKAASAVIRKTLDPDAQIKAFYKHLTR-----TDYDNPALVIAAGYAAANLPSAM 434
QY 422 ARAVARETIIKQYQTYWQEBDKAVNLKMRDPIFEEMTEPEPGVPDNTQVESKGF 481
DB 435 ATATVARDVSIKWLFWNLWATEDEDRNLAWREFFRDVFAATGVRPS-----GVN 487
QY 482 ECGEYFNPVDLANN--WKNKGYGALFLYFLGNLRLIKAKMLDPNEIFTNKOSI 534
DB 488 DGAFTNYADADLADPALNTSGIGMNTLYKDGICRLQAAKTQWDPRNVFTHALGI 542

RESULT 7

Q8VMA5 PRELIMINARY; PRT; 489 AA.
ID 08VMA5; 08VMA5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ACLO.
GN ACLO.
OS Streptomyces galilaus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=33899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3AR-33;
RA Chung J., Fujii Y., Tsukamoto N., Sankawa U., Ebizuka Y.;
RT "Aklavinone-acylchomycin biosynthesis gene cluster from Streptomyces galilaus";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008446; BAB7054.1; -
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
SQ SEQUENCE 489 AA; 53652 MW; BA006ADA0BA33E7 CRC64;

Query Match 19.7%; Score 590.5; DB 2; Length 489;
Best Local Similarity 29.3%; Pred. No. 3.6e-37; Mismatches 204; Indels 89; Gaps 17;
Matches 159; Conservative 91;

QY 22 DKPRLPSM-KGFNRWIGTINIDFVVVYTPQAGCTALDRAMEKSPGTIVRIVSGGHC 80
DB 2 DBADRRYODLVIRGFGFRG- RPDVYVYVHTADQVDAVNAQV--AAGRRIAVNSGHC 58
QY 81 YEDPVPDECVAIINVTGLVSGYDD-DRGYVSSGDTNMGSKTLFRDHGVLFGSCY 139
DB 59 FEGFVDDPAVRAVIDMSQWRQVYDSGKRAFAVEPGATLGETYRALYLDMGVTIPAGVCP 118
QY 140 SYGLGGHIVGGGDLARLHGLPVDMLSGVEVYVPLTEDSVLKYVHDSGND--GEL 197
DB 119 QGVGGHVLGGGGLPSRRDGVADHLVAEVV--VVDASQARKVVA TSAADPNREL 175
QY 198 FMAHTGGGNGFIITKYFKD-----LWSPRGVIA SNLHFSMDGFTRDALQ 246
DB 176 WMAHTGGGNGFIITKYFKD-----LWSPRGVIA SNLHFSMDGFTRDALQ 246
QY 247 LITKY-----FKLARCWKNTVKGFOIFHOAAEEFVYLYTSYNDAREVAQDRHYL 300
DB 236 IIDNKGAMHQRNSAAGTPYASMSVYILNSRAAGQILDLIDIGDGLDGAETLLND----F 291
QY 301 EADHQ-----IKTCGPTKALGSHAGMARPPVPRKRHNSKTSYMDENMDYFYAL 353
DB 292 VAAVEGTCVPAVQRTPE-----WL-----RATLANKF----- 321
QY 354 TETINGSGPNRGKYSAYMIDFPDFOIDVIMKYLTVEPDLGTSAMKDALLOVDMFG 413
DB 322 -----DTGDFRDKSGALRKRPWTAQAATLYRHLS--ADSQWCE-----VSLYSYG 369
QY 414 EHKVVMATYVAQREYIIKQYQTYWQEBDKAVNLKMRDPIFEEMTEPEPGVPDNTQ 473
DB 370 KNSVPEATATVARDVSIKWLFWNLWATEDEDRNLAWREFFRDVFAATGVRPS-----GVN 487

QY 474 VESGKGVFEGCYENYPDVLAN--WKNKGYGALFLYFLGNLRLIKAKMLDPNEIFTNK 531
DB 430 T-----EGTINFPDVLADPRMNTSGVPMYITLYKGNVRLQVYARAMPDPRVFRIA 482
QY 532 OSI 534
DB 483 LSV 485

RESULT 8

Q9ZAR8 PRELIMINARY; PRT; 485 AA.
ID 09ZAR8;
AC 09ZAR8;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DNRW.
GN DNRW.
OS Streptomyces peucetius subsp. caesiuss.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=55158;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27952;
RA Hong Y.-S., Hwang C.K., Lee J.J.;
RT "Regulation of anthracycline production in Streptomyces peucetius by the dnrW";
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U80222; AAD00354.1; -
DR InterPro; IPR001899; Gram pos anchor.
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.
SQ SEQUENCE 485 AA; 52227 MW; 07DC310677034557 CRC64;

Query Match 14.7%; Score 440.5; DB 2; Length 485;
Best Local Similarity 28.3%; Pred. No. 1.3e-25; Mismatches 183; Indels 79; Gaps 15;
Matches 135; Conservative 80;

QY 25 DPLRPSM-KGFNRWIGTINIDFVVVYTPQAGCTALDRAMEKSPGTIVRIVSGGHC 83
DB 49 DPRYDIVGRGINAFT--PDPTVAVATAEQAQVDSVADGRILAVR--TGHCHEFS 105
QY 84 FVFDCEVAIINVTGLVSGYDDRGYF-VSSGDTNMGSKTLFRDHGVLFGSCYSG 142
DB 106 LVDDPAVTVVIDVSEMRGVEYFDEGLNAFVSAGATLGTMYRSLYLGMVTVPAGRCPYVG 165
QY 143 LGGHIVGGGDLARLHGLPVDMLSGVEVYVPLTEDSVLKYVHDSGND--GELFWA 200
DB 166 VGGHIAAGGGGALSTYGLSDHLHGEV--VVDSGRAVLVATREPSDPHRLDWWA 222
QY 201 HTGGGNGFIITKYFKD-----LWSPRGVIA SNLHFSMDGFTRDALQ 245
DB 223 HTGGGNGFIITKYFKD-----LWSPRGVIA SNLHFSMDGFTRDALQ 245
QY 246 DILTYFKLARCDWKNTVKGFOIFHO-----AAEFVYLYTSYNDAREVAQDR 296
DB 283 TL-----RNFETWHERHAAPDGPARDNSIALPRGTGGPPLTETAVIDA 327
QY 297 HYHLEADIEQIKTC--EPTKALGSHAGMARPPVPRKRHNSKTSYMDENMDYFYAL 354
DB 328 ---TRPDABELTDRITREVSRSVG-----APPEVSVT-----LFWLAT 364
QY 355 ETINGSGPNRGKYSAYMIDFPDFOIDVIMKYLTVEPDLGTSAMKDALLOVDMFG 413
DB 365 LVPDEFAGIKGKFSKAFRLRGWSEQAQRMVYQRLTTSQVHNAA--ATVYLLSHSG 420
QY 414 EHKVVMATYVAQREYIIKQYQTYWQEBDKAVNLKMRDPIFEEMTEPEPGVPDNTQ 473
DB 421 EVNRPADDTMAHRAVATLKYWSVFWFDEREDLALHDMVNSYAFEGDAGGVDP 477

RESULT 9
P71091 PRELIMINARY; PRT; 480 AA.
ID P71091
AC P71091
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, last annotation update)
DE Hypothetical 54.4 kDa protein (YGAK protein).
GN YGAK
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Cummings N.J., Connerton I.F.,
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 114-480 FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogatawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,
RA Bortner R., Bortner L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA Ertan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Fuma S., Gallizi A., Galleon N.,
RA Gilm S.Y., Glaeser P., Goffeau A., Goldrich E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasaara Y., Klier-Richard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano C.,
RA Kurita K., Lapidus A., Lardinois M., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle B., Potwolik S., Prescott A.M.,
RA Pressecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis.";
RT Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 114-480 FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogatawara N., Yoshikawa H., Danchin A.,
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z82044; CAB04812.1; -
DR EMBL; Z99108; CAB12708.1; -
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
KM Complete proteome.
SQ SEQUENCE 480 AA; 54477 MW; B43CD118F7F4C8B CRC64;
Query Match 13.3%; Score 399; DB 16; Length 480;
Best Local Similarity 26.0%; Pred. No. 2e-22;
Matches 144; Conservative 66; Mismatches 192; Indels 152; Gaps 21;
QY 20 TPDKPDPLPSMGKGFNRKRGITNI-----DFYVVYVTPGACTALDRAMEKKSPTGVR 73
DB 41 TRDDPD-----YNEARTNINLSIERYPDIIVFCNKQDALNALKWARENRVP--FR 89

QY 74 IVSGGHCYEDF-----VFD--ECVKAIINVTGLVSGYDDRGYFVSSGDTNNGSEFT 124
DB 90 INGGHRSYENFNLNNGVLIDISEMKKITVN-----QPKLAIYEAGAELEGEVRYT 140
QY 125 LRRDHGRVLPGGSCSYVIGGHIIVGGGDCIILARLHGLPVDMLSGVEVYVKKPVLTDSDVK 184
DB 141 LMQ-YGLTLPAQITANVITGLTLLGGIGLLTFRAGLTCTDSIVOLEMTVA----- 189
QY 185 YVHKDSEG-----NDGELEFMAHTGGGGNFGIITKYFKPD.PMSPRGVIAISLHPS 235
DB 190 ---DEKEGDLITVSGSNHPL.FMASGGGGNFGIIVISMTKAPISGVSLF--SITMG 244
QY 236 WDGFTRDALQDLLTKYFKLARCWMKNTYKFOI.FHOAAEFVWYLYTYSYNDAREVAD 295
DB 245 WDFEE-----VYNTWQN-----WPTYTD----- 264
QY 296 RHYHLEADLEQIYKCEPTKALGHAAGAPF-----VRPRKHTSTSYMHDETMXY 348
DB 265 ---RLTSSIEFWPEKVNRIEALGQFVG---PKTELKLLKPLKLAGSPSGVKT--- 314
QY 349 PFYALTRTINGSGPNQGRKY--SAVMIKDPDFQIDVIMKYLTEVPDGLTSAEMKALL 406
DB 315 PFLVATFNSFGNDPQMKRSGSPTIEKLEBRALSTIKHLEHAPN-----QNASV 367
QY 407 QVDMFGGEIHKYVWDATAVAOREYIIKQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 466
DB 368 WQALGGAAGRAVAPDQTAIFYRDALIAQEYLITNMTSPGKRONVWIEGLRTSLSK--- 424
QY 467 VDPENQVSGKGVGECGCTFNYPDVDLNNKNGKCALELYLGNLRIKAKMLMDPNE 526
DB 425 -----TWGDYVNWPDIEIRNWPRTYGE-----NVERLRVKTYYTDPEN 463
QY 527 IFNKOSIPTKPLX 540
DB 464 VRFEGSIP--PLR 475
RESULT 10
Q9FZC5 PRELIMINARY; PRT; 530 AA.
ID Q9FZC5
AC Q9FZC5
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, last annotation update)
DE TIK7.23 protein.
GN TIK7.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Sakano H., Liu S.X., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA Chin C., Chiu J., Choi E., Chung M., Gonzalez A., Howing B., Liu A.,
RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA Conway A.B., Hansen M.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA Miranda M., Nguyen M., Palm C.J., Shim P., Southwick A., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.,
RT "The sequence of BAC TIK7 from Arabidopsis thaliana chromosome 1.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC013427; AAP98577.1; -
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
SQ SEQUENCE 530 AA; 58947 MW; 116C6CA06B4C322 CRC64;
Query Match 12.3%; Score 368.5; DB 10; Length 530;
Best Local Similarity 27.0%; Pred. No. 5.3e-20;
Matches 134; Conservative 76; Mismatches 180; Indels 107; Gaps 23;
QY 72 VRIVSGHCYEDFVDECVKAI-----NVTGLVSGYDDRGYFVSSGDTNNGSFKT 124

Db 103 LRIRSGHDEGLSYSSVFPVILDMYNLRSLTVDS-----KKAMIOAGAT-LGELYT 156
QY 125 LFRDGRV--PGGSCYSVGLGHIIVGGDGLIARHGLPVDMLSGVEVNVKVLTEDSV 182
Db 157 NNNDVSQTLAPFAGVCAITAGAGHISGGYGIMLRKYGITVHVDAQII-----DVA 209
QY 183 LKVVHDSGNDGELFWAHTGGGNGFIIT--KYFKDLPMSPRGVIANLHFSWDGFT 240
Db 210 GLILNRAITMGED--LFWALRGGGGSGFVILSMKINLVDPKIVT-VFKNVNTLEGG-- 264
QY 241 RDLADLLTKYFKLARCWKNTVGF--QIFHOAAE-----FVMTLYTSYSD 287
Db 265 ----TDVLYK-----WQVASKPEPSLFVRAMPQVANGTRGGRITTVFVYQFLGR 312
QY 288 AEREVAQDHHYHLEADIEQIYKTCERTKALGHAAGAPFPVRPRKHTSKTSYMHDETMD 347
Db 313 TDLMAIMNQNPDELGLK--HEDCOEWSMLNSTLFWADYPA-----GPTSTL----LD 360
QY 348 YPFYALTETINGSQPNQSGKYSAMIKDPFDIVIKVLTVEPDGLTSAEMKDALLQ 407
Db 361 RP-----SSPDFFKSKSDYVKRPIKEGLEKMTKMKRNNT-----VMNQ 403
QY 408 VDMFGEIHKVVDATAVAQRE-YIIKIQYQYWOEDKDAVNLKMRDPEMEYEGG 466
Db 404 FMYGGMVRIPATATAFPHRKGNLFKIQYFTTFWANAATMSLSQMKELY-EVABPY-- 460
QY 467 VDPDNTQVSSGKGVFGCVFNPVDVNLNWKNG-----YGALELYFLGNLRLIA 518
Db 461 -----VSSNP--REAFNRYRIDVGSNPSGETVNDVDEAKIYGS--KYFLGNLKRLLMDV 508
QY 519 KWLMDPNEIFTNKOSIP 535
Db 509 KAKYDPDNFKNQOSIP 525

RESULT 11
Q949N1
ID 0949N1 PRELIMINARY; PRT; 530 AA.
AC 0949N1;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Hypothetical 58.9 kDa protein.
GN T1K7.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh U., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toritani M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Kosemura E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.,
RA Sakurai T., Satou M., Seki M., Shin P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full Length cDNA of gene T1K7.23 (GI:9797759)."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051000; AAK93677.1; -
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
KW Hypothetical protein.
SQ SEQUENCE 530 AA; 58947 MW; 1BC6C456BDE9A3F4 CRC64;

Query Match 12.3%; Score 368.5; DB 10; Length 530;
Best Local Similarity 27.0%; Pred. No. 5.3e-20;
Matches 134; Conservative 76; Mismatches 180; Indels 107; Gaps 23;
QY 72 VRIVSGHCEYDFEDECVAII-----NVTGLVESGYDDDRGYFVSSGDTNMGSPKT 124

Db 103 LRIRSGHDEGLSYSSVFPVILDMYNLRSLTVDS-----KKAMIOAGAT-LGELYT 156
QY 125 LFRDGRV--PGGSCYSVGLGHIIVGGDGLIARHGLPVDMLSGVEVNVKVLTEDSV 182
Db 157 NNNDVSQTLAPFAGVCAITAGAGHISGGYGIMLRKYGITVHVDAQII-----DVA 209
QY 183 LKVVHDSGNDGELFWAHTGGGNGFIIT--KYFKDLPMSPRGVIANLHFSWDGFT 240
Db 210 GLILNRAITMGED--LFWALRGGGGSGFVILSMKINLVDPKIVT-VFKNVNTLEGG-- 264
QY 241 RDLADLLTKYFKLARCWKNTVGF--QIFHOAAE-----FVMTLYTSYSD 287
Db 265 ----TDVLYK-----WQVASKPEPSLFVRAMPQVANGTRGGRITTVFVYQFLGR 312
QY 288 AEREVAQDHHYHLEADIEQIYKTCERTKALGHAAGAPFPVRPRKHTSKTSYMHDETMD 347
Db 313 TDLMAIMNQNPDELGLK--HEDCOEWSMLNSTLFWADYPA-----GPTSTL----LD 360
QY 348 YPFYALTETINGSQPNQSGKYSAMIKDPFDIVIKVLTVEPDGLTSAEMKDALLQ 407
Db 361 RP-----SSPDFFKSKSDYVKRPIKEGLEKMTKMKRNNT-----VMNQ 403
QY 408 VDMFGEIHKVVDATAVAQRE-YIIKIQYQYWOEDKDAVNLKMRDPEMEYEGG 466
Db 404 FMYGGMVRIPATATAFPHRKGNLFKIQYFTTFWANAATMSLSQMKELY-EVABPY-- 460
QY 467 VDPDNTQVSSGKGVFGCVFNPVDVNLNWKNG-----YGALELYFLGNLRLIA 518
Db 461 -----VSSNP--REAFNRYRIDVGSNPSGETVNDVDEAKIYGS--KYFLGNLKRLLMDV 508
QY 519 KWLMDPNEIFTNKOSIP 535
Db 509 KAKYDPDNFKNQOSIP 525

RESULT 12
Q9SVG3
ID 09SVG3 PRELIMINARY; PRT; 539 AA.
AC 09SVG3;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Reticulin oxidase-like protein.
GN F21C20.190 OR AT4G20840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Mayer M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Beyer K.F.X., Lemcke K., Schueller C.,
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL080254; CAB45850.1; -
DR EMBL; AL161553; CAB79084.1; -
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
SQ SEQUENCE 539 AA; 60143 MW; 6645E9376E216426 CRC64;

Query Match 11.6%; Score 347.5; DB 10; Length 539;
 Best Local Similarity 26.2%; Pred. No. 2.3e-18;
 Matches 141; Conservative 76; Mismatches 181; Indels 141; Gaps 28;

QY 52 TPQACATLDR-----AMKCSFG---TYRIVSGHCYE--DFVPECKAIIINVTGLV 100
 DB 83 TPKPAIIVTPRSDIHVSAATVCSKSNFLKIRSGHDEGLSTISDKPF--FLIDMSNLR 141
 QY 101 ESGVD--DRCGYFVSSGDT-----NMGSFKTLFRDGRVLPGGSCSYVGLGHHVGGD 152
 DB 142 DVSDVIDDQASAMISAGATLGEVYRIKESKV---HG--FPAGVCTPVGGHISGGY 135
 QY 153 GIARLHGLPYDMLSGVEVVKPVLTEDSVLKYVHKDSEGNDGLFMAHTGGGSGNGIIT 212
 DB 196 GNMRLKFGLSVDNLIDAKIV-----DVNGQILDRKSMGSD--LFWAISGGGASPGV 246
 QY 213 TKYFKDLPMSPRVIT-----ASNLFHSDG-----FTRDLQDILLTYFK 253
 DB 247 LGYKVKLVPVETVTRVEKYMDSGAVDMVHKQSVGPKTRNLFLRMLIQPTRRKKV 306
 QY 254 LARGDMKNTGKFOIFQAAEEFYVNYLTSYSDAREVAQDRHYHLEADIEQIYKTCER 313
 DB 307 TVRA---TV--VALFGRAEVVALIGKEFP-----ELSLKK--ENCSE 343
 QY 314 TKALGHAGW---APPPVPRKRHTSKTSYMHDETDYFPYALTETINGSQPNQSKYK 369
 DB 344 MTWFGSALMMDNRVNPFOIDK-----VFLERLNDRANF-----GKRX 381
 QY 370 SAWIKPFPFQIDIVIKYLTEVPD--GLTSAEMKDALIQDMFGGELHKVWDATAVAQR 428
 DB 382 SDVYASISIPRDISLFRKMTLQKIGLV-----FNPYGGKAEVTVNATPPPHR 431
 QY 429 EYIKLQYQTVQWEDDAVNLKWRDPFYE--MYEPYGGVDPPTQVESGKGVFEGCY 485
 DB 432 SKLFKIQYSVTWQ--NSVEIE--KGLFQANVLYSFMGFSVSKNR-----NAY 477
 QY 486 FNYPDVVL-----NNMKNGK--YGALELYFLGNLRLIKAKWIMDNEIIFTNKOSIPT 536
 DB 478 LNYRDVIDIGVNDHGTNSYEGEVYG--RKYFGDNFRLVKVTAADPNFNRNQSIP 534

RESULT 13
 Q9SVG7 PRELIMINARY; PRT; 528 AA.
 AC Q9SVG7;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Hypothetical 58.8 kDa protein.
 GN F21C20.150 OR A74G20800.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Beyan M., Pohl T., Weizengger T., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueler C.;
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Pohl T., Weizengger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL080254; CAB45846.1; -.

DR EMBL; AL161553; CAB79080.1; -.
 DR Interpro; IPR001575; Oxid_FAD_bind.
 DR Pfam; PF01565; FAD_binding_4; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 528 AA; 58846 MW; D70E7667AABAE3E7D CRC64;

Query Match 11.5%; Score 346; DB 10; Length 528;
 Best Local Similarity 27.0%; Pred. No. 2.9e-18;
 Matches 134; Conservative 68; Mismatches 187; Indels 108; Gaps 23;

QY 72 VRTVSGHCYEDFVPECKA--AIINVTGLVSGYDDDRGY--FVSSGDTMGSKTLFRH 129
 DB 102 ITRISGHDYEGISFTSSVPVILDMHDLRSITIDFRQAWVDGAT--MGLYTKIAAA 160
 QY 130 GRVL--PGSSCYSVGLGHHVGGSDGILLARLHGLPYDMLSGVEVVKPVLTEDSVLKYVH 187
 DB 161 SKTLAFAGVCPTLGGHISGGGYGNLIRKXGISDVHVDARIV----- 205
 QY 188 KQSEGN-----DGLFMAHTGGGSGNGIIT--KYFKDLPMS-----PRGYI 228
 DB 206 -DVNGNLTGATLGRDLWALIRGGGASFGVILSWKINLVDPKTVTRKVKLTLEQGV- 263
 QY 229 ASNLHFSWQFTRDALQDILLTYFKFLARCDMKNTVCKFOIFQAAEEFYVNYLTSYSDA 288
 DB 264 -TDVLKQVLVSKLPQDIFLRAMP-----KPVNGV---PSEKTIIVVFYAPLGS 312
 QY 289 EREVAQDRHYHLEADIEQIYKTCPTKALGHAGWAPFPVRPRKRHTSKTSYMHDETDY 348
 DB 313 RLMAIMAKNLPBLGAKR--EDCYEMSWINTTTFMQNYV-----GTSYVL--LDR 360
 QY 349 PFYALTETINGSQ--NQRKYSAYMIKDPDFQIDIVIKYLTEVPDGLTSAEMKDALIQ 407
 DB 361 F-----SGPAAPAFYKSKSDYVKRPIPKXEMEKIKWAM-----LKFNNMNO 401
 QY 408 VDMFGEHKKVWDATAVAVAQRE--YIIKLQYQTVQWEDDAVNLKWRDPFYEEMYPYGG 466
 DB 402 WNYGGMVDLIPADATAPRRKGNLFKIQYFALMTDANATYANGLMROIYHEM--EPY-- 458
 QY 467 VPDPTQVESGKGVFEGCYFNYPDVLDNNMKNGK-----YGALELYFLGNLRLIK 518
 DB 459 -----VSSNP---REAFINLRDIDVGSNPSGETLLEBAKTYGS--KYFLGNFKRLMEV 506

QY 519 KWLMDNEIIFTNKOSIP 535
 DB 507 KAKYDDENFREFQOSIP 523

RESULT 14
 Q9AYM8 PRELIMINARY; PRT; 535 AA.
 AC Q9AYM8;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE CPRD2 protein.
 GN CPRD2.
 OS Vigna unguiculata (Cowpea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolaeae; Vigna.
 ON NCBI_TaxID=3917;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Iuchi S.;
 RL "Drought inducible gene from cowpea."
 DR EMBL; AB056448; BAB33033.1; -.
 DR Interpro; IPR001575; Oxid_FAD_bind.
 DR Pfam; PF01565; FAD_binding_4; 1.
 SQ SEQUENCE 535 AA; 60291 MW; 7F889FCBF52FD80B CRC64;

Query Match 11.5%; Score 346; DB 10; Length 535;
 Best Local Similarity 26.2%; Pred. No. 2.9e-18;

Matches 138; Conservative 70; Mismatches 207; Indels 112; Gaps 24;

QY 49 VVYTPGACATLADRAMKSPG---TVRIYSGHCYEDFVDECYK-AIINVTGLVSGY 104
 Db 78 VVTPG--VSHIQAIIKCSQRHGLQIRIRSGHDYEGSLYARVFLIDLINFEIKY 135
 QY 105 D-DDRGYVSSGDTNWGSFETLFRDHGRV-LPGSGCYVGLGHIYVGGDGLIARLHGP 162
 Db 136 DVENRTAMVQVATLDELTYTISQASKTLGFPAGVCYVAGGHSIGGGYGLMRKYGLA 195
 QY 163 VVMLSGVEVYVAVLTEDSVLKVHDSBN-----DGELFMAHTGGCGGNGFIITKX 215
 Db 196 ANNVIDAHII-----DVGNNLDRKMGEDLFMALRGGGGASFGYIVSM 239
 QY 216 YFADLPMSRGVI-----ASNLFHSMDFGTALQDILLTKYFKLARCWMKTVK 265
 Db 240 KTKLVFVSTVYFVFNVERLLENATEIIEKMQ-LVANKLDERIFLMDLARAN-SSQHG 297
 QY 266 FQI-----FHQAEFVWYLYTSYNDAREVAODRHYHLEADIEQIYKTCPTKALG 318
 Db 298 LALQANFVAMFGGVEBELIPMQKNP-----ELGLK-----KOCETSWIG 340
 QY 319 GHAGMAPPVPRPKRHTSKTSYMHDETMDFPALTEITNGSGPNRGKYKAYMIKDP 378
 Db 341 S-----AVF-----TNGALIGSSGHEAPVTLNRTQIRSG--KYGKSDYVRRKPI 384
 QY 379 DFQIDVIMKYLTPEPDGLTSAEMKDALLOVDFMGGEIHVKVMDATAVAORE-YIILQYQ 437
 Db 385 VGLRLKMLWLD-----DKVYSOLOPARYGKMDNISESEIIPAHSGYIIFHIHY 437
 QY 438 TYWBEEDKDAV-NLKMIRDFEEMEPYGGVDPNTQVESGKVEGCFNYPDVLLNN 495
 Db 438 VVWQEGDEATQHVWMIIRLYKYM-EPY-----VSNSP---RAAYVNRDIDIG 484
 QY 496 WKNG-KYGALE----YFLGNLRLIKAKMLDPEIFPNKOSIP 536
 Db 485 NNNGYTSYHQASIMGLKFSNNFKRLATYKVDPHNPFNEQSIPT 531

RESULT 15

Q9FZC8 PRELIMINARY; PRT; 529 AA.

AC Q9FZC8; 01-MAR-2001 (T-EMBLrel. 16, Created)

DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)

DE TIK7.20 protein.

GN TIK7.20.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxId=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Sakano H., Liu S.X., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,

RA Chin C., Chiu J., Choi E., Chung M., Gonzalez A., Hwang B., Liu A.,

RA Vaysberg M., Altati H., Brooks S., Buehler E., Chao Q., Conn L.,

RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,

RA Miranda M., Nguyen M., Palm C.J., Shin P., Southwick A., Davis R.W.,

RA Ecker J.R., Federspiel N.A., Theologis A.;

RT "The sequence of BAC TIK7 from Arabidopsis thaliana chromosome 1.";

RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC013427; AAF98574.1; -

DR InterPro; IPR001575; Oxid_FAD_bind.

DR Pfam; PF01565; FAD binding 4; 1.

SQ SEQUENCE 529 AA; 59217 MW; 0DC786AA3CB0590D CRC64;

Query Match 11.1%; Score 333.5; DB 10; Length 529;

Best Local Similarity 25.8%; Pred. No. 2.6e-17;

Matches 128; Conservative 77; Mismatches 184; Indels 107; Gaps 22;

QY 72 VRIYSGHCYEDFVDECYKALI-----NVTGLVSGYDDRGYVSSGDTNWGSFET 124
 Db 103 LIRSGGHDLEGLSYRSSVPFYLDMFNLRSLITVNLV-----KRAMVQAGAT-LGELYV 156
 QY 125 LPRDHGRV--FGGCTSYVGLGHIYVGGDGLIARLHGLPVMLSGVEVYVAVLTEDSV 182
 Db 157 KINEASQTLAFAGVCPYVGVGHIISGGYGNLMKRFGIIVHVSDAOI-----DVN 209
 QY 183 LKYVHDSSENGELFMAHTGGGNGNFGIITKYFPKDLPMSPR--GVLSNLFHSMDFPT 240
 Db 210 GKLINRNASGED-LFMALRGGGGASFGVILSWKI-NLVKVKILTVFKVNTLEGG-- 264
 QY 241 RDAQDLTKYFKLARCWMKTVGRF--QIFHOAAEEFV-----MYLYTSYND 287
 Db 265 -----TDVLK-----WQLVATKFPEDLFWRAMPQIINGARGDRTIAVVFYQFLGP 312
 QY 288 ARREVAQDRHYHLEADIEQIYKTCPTKALGHAGMAPPVPRPKRHTSKTSYMHDETM 347
 Db 313 ADKLALINMQRLPEIGLR--EDCHEMSWFNTTLFMADYPAGTPK-----SVLLDRPTN 364
 QY 348 YFFYALTEITNGSGPNRGKYKAYMIKDPDFQIDVIMKYLTPEPDGLTSAEMKDAL-L 406
 Db 365 PGFP-----KSKSDYVKKPIPREGLEKLMK-----TMFKFNIVVM 400
 QY 407 QVDMFGEIHKVMDATAVAORE-YIILQYQTYWQEBDKDAVNLKMIRDFEEMEPY 465
 Db 401 QFNPYGVVMDQIPSTATAPPHKGMFKVQYSTTWLANATEISLSMKEL-KVAEPY- 458
 QY 466 GVPDPNTQVESGKVEGCFNYPDVLLNNMKNGKGALE----YFLGNLRLIKAK 519
 Db 459 -----VSSNP---REAFFNVRDIDIGSNPSDETNDVDAIKYGYKFLGNLKLIMQYK 507
 QY 520 WLMDEIFETNKOSIP 535
 Db 508 AKYDPENFPKNOSIP 523

Search completed: June 10, 2003, 09:59:59

Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:49:56 ; Search time 24 Seconds

(without alignments)
943.586 Million cell updates/sec

Title: US-09-998-284-2

Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVDVNAGT.....IFTNKQSIPTKPKKPKQTK 546

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	290.5	9.7	535	1	RETO_PAPSO
2	286.5	9.6	538	1	RETO_PAPSO
3	154.5	5.2	447	1	MCRA_STRLA
4	118.5	4.0	769	1	DIVL_CAUCR
5	116	3.9	478	1	MM03_PABIT
6	111	3.7	1009	1	YE68_METUA
7	109.5	3.7	496	1	DLU3_YEAST
8	104	3.5	661	1	YJCS_ECOLI
9	103	3.4	634	1	ELM1_ASPPU
10	102	3.4	634	1	ELM2_ASPPU
11	102	3.4	702	1	AMYG_CLOSP
12	101.5	3.4	969	1	DPMW_NEUTN
13	97.5	3.3	836	1	MGEL_SFVKA
14	97	3.2	469	1	MM01_BOVIN
15	97	3.2	955	1	AGAB_VIRB7
16	95.5	3.2	530	1	ALP2_YEAST
17	95	3.2	847	1	AGUA_TIRRE
18	94.5	3.2	359	1	RF2_BORBU
19	94.5	3.2	494	1	KICL_HUMAN
20	94.5	3.2	1301	1	PTP9_DROME
21	94	3.1	280	1	SPEE_AOUAE
22	94	3.1	388	1	CMF2_SCHPO
23	93.5	3.1	562	1	EST1_CABEL
24	92.5	3.1	290	1	HTRL_ECOLI
25	92.5	3.1	1287	1	DCOR_CHICK
26	92	3.1	1287	1	RPO1_FOMPV
27	92	3.1	4655	1	LRP2_HUMAN
28	91.5	3.1	397	1	PAPS_BACSU
29	91.5	3.1	458	1	HDNO_ARTOX
30	91.5	3.1	663	1	MM02_CHICK
31	91.5	3.1	1290	1	BXB_CLOBO
32	91	3.0	787	1	K6PF_DROME
33	90.5	3.0	477	1	MM03_MOUSE

34	90.5	3.0	534	1	FM02_HUMAN
35	90.5	3.0	559	1	HUTV_BACHD
36	90.5	3.0	1263	1	RPOB_THEMA
37	90	3.0	477	1	NIRD_METMP
38	90	3.0	946	1	K6P2_CANAL
39	90	3.0	990	1	K6P1_PICPA
40	89.5	3.0	467	1	MM08_HUMAN
41	89.5	3.0	514	1	SYS_METMP
42	89.5	3.0	1006	1	EGAL_ASPPG
43	89.5	3.0	2266	1	TOXB_CLODI
44	89	3.0	680	1	GAOX_DACDE
45	89	3.0	987	1	K6P1_YEAST

ALIGNMENTS

RESULT 1	ID	RETO_PAPSO	STANDARD;	PRT;	535 AA.
AC	P93479;				
DT	15-JUN-2002	(Rel. 41, Created)			
DT	15-JUN-2002	(Rel. 41, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Reticuline oxidase precursor (EC 1.5.3.9) (Berberine-bridge-forming enzyme) (BBE) (Tetrahydroprotoberberine synthase).				
GN	BBE1.				
OS	Papaver somniferum (Opium poppy).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;				
OC	Papaveraceae; Papaver.				
OX	NCBI_TaxID=3469;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Marianne;				
RX	MEDLINE=9127853; PubMed=8972604;				
RA	Facchini P.J., Penzes C., Johnson A.G., Bull D.;				
RT	"Molecular characterization of berberine bridge enzyme genes from				
RT	Opium poppy"; 112:1669-1677(1996).				
RL	Plant Physiol. 112:1669-1677(1996).				
CC	- FUNCTION: Essential to the formation of benzophenanthridine				
CC	alkaloids in the response of plants to pathogenic attack.				
CC	Catalyzes the stereospecific conversion of the N-methyl moiety of				
CC	(S)-reticuline into the berberine bridge carbon of (S)-scoulerine.				
CC	- CATALYTIC ACTIVITY: (S)-reticuline + O(2) = (S)-scoulerine +				
CC	H(2)O(2).				
CC	- COFACTOR: FAD AND METAL ION.				
CC	- PATHWAY: Benzophenanthridine alkaloids biosynthesis.				
CC	- SUBCELLULAR LOCATION: VESICULAR.				
CC	- SIMILARITY: Belongs to the oxygen-dependent FAD-linked				
CC	oxidoreductase family.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AF025430; AAC61839.1; .				
DR	InterPro; IPR001575; Oxid_FAD_bind.				
DR	Pfam; PF01565; FAD_binding_4.1.				
DR	PROSITE; PS00862; OX2_COVAL_FAD; 1.				
KW	Oxidoreductase; Signal; Glycoprotein; Flavoprotein; FAD;				
KW	Alkaloid metabolism.				
FT	SIGNAL	1	23	POTENTIAL.	
FT	CHAIN	24	535	RETICULINE OXIDASE.	
FT	BINDING	108	108	FAD (COVALENT) (BY SIMILARITY).	
FT	CARBOHYD	42	42	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	475	475	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SQ	SEQUENCE	535 AA;	59903 MW;	F0341EF38AB41239 CRC64;	

Query Match 9.7%; Score 290.5; DB 1; Length 535;
 Best Local Similarity 23.8%; Pred. No. 7.5e-15;
 Matches 123; Conservative 87; Mismatches 160; Indels 147; Gaps 28;

QY 71 TVRIVSGGCHCE-----DFVDECVKA-----IINVTGLVSGYDDRGFVSSGD 116
 DB 100 TIRLRSGGHSYEGLSYADPPFVIVDMNINRISIDVSEFVAVWESATLGEIYVAIAQS 159
 QY 117 TTMSSFTLFRDHRVLPGGSCSYVGLGHIYVGGDGLIARHGLPVDWISGVEVVPV 176
 DB 160 TDTLGF-----TAGWCPTVSGSGHISGGGFGFMSKRYGLAAD-----NVDAI 202
 QY 177 LTEDSVLKRYHKGSENGDELFWAHTGGGNGFIITKRYFKD.PMSPRGV----- 228
 DB 203 LI-DSNGAILDRBKMGDD--VFMAIRGGGGVWGAIVAMKIKLIPVEKTLVFRVTKVNG 259
 QY 229 ---ASNHFMSWDGTRDALQDLTKYFKLACDKMKNVKGQIHOAAEEVWMLYTSYS 285
 DB 260 IEDASSLHKWQ-YVADELDEDF-----VSVLGV-----NGNDAMLFLGLHLG 304
 QY 286 -NDAREVAODRHYL---EADIQIYKTCPTKALGGHAGMAPFVPRKRTSKTSYM 341
 DB 305 RKDAKTIIDKPELGLVDEFEPM-----SWG----- 333
 QY 342 HDETMDPYVALTETTINGSGPNRGKY-KSAVMK-DPPDFQIDV-IMKLTLEVPDGLTS 398
 DB 334 --ESM--AFISGLDITSELN-NRFLKDERAFKTKVDFTKVSPLVFNRLHEW-----L 383
 QY 399 AEMDALLQVDFEGEHLKVMVDTAAVQ-----EYIKLOYOTYNOBEDDAVN-- 449
 DB 384 SEQGFIALNGFGKMSSEISTDTPPHRKTKLMEFYIA-----MMODESKJGEF 437
 QY 450 LKMIKRDYFEEWYDYGVPDNTQVSGKGVFEGCYFNPVDLN--NMKNY--GKYGAL 505
 DB 438 SEMLAKEYDYI-----EPFVSKERPVG-----YNNHIDLDIGIDMKRKSSTYNAVE 484
 QY 506 L-----YFLGNLNLRIKAKMLMDPNEIFTNKQSIIP 535
 DB 485 IARNWGERYFSSNYERLVKAKTLIDPNNVFNHPQISIP 521

RESULT 2
 RETO_ESCCA STANDARD; PRT; 538 AA.
 ID RETO_ESCCA
 AC P30985;
 DT 01-JUN-1993 (Rel. 26, Last Created)
 DT 01-JUN-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reticuline oxidase precursor (EC 1.5.3.9) (Berberine-bridge-forming enzyme) (BBE) (Tetrahydroprotoberberine synthase).
 GN BBE1.
 OS Eschscholzia californica (California poppy).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
 OC Papaveraceae; Eschscholziaceae; Eschscholzia.
 OC NCBI_Taxid=3467;
 RP [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=92052284; PubMed=1946465;
 RA Dittich H., Kutchan T.M.;
 RT "Molecular cloning, expression, and induction of berberine bridge
 RT alkamide, an enzyme essential to the formation of benzophenanthridine
 RT Proc. Natl. Acad. Sci. U.S.A. 88:9969-9973(1991).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98145481; PubMed=9484487;
 RA Hauschild K., Pauli H.H., Kutchan T.M.;
 RT "Isolation and analysis of a gene bcl encoding the berberine bridge
 RT enzyme from the California poppy Eschscholzia californica.";
 RL Plant Mol. Biol. 36:473-478(1998).
 CC -FUNCTION: Essential to the formation of benzophenanthridine
 CC alkaloids in the response of plants to pathogenic attack.

CC Catalyzes the stereospecific conversion of the N-methyl moiety of
 CC (S)-reticuline into the berberine bridge carbon of (S)-scoulerine.
 CC -CATALYTIC ACTIVITY: (S)-reticuline + O(2) = (S)-scoulerine +
 CC H(2)O(2).
 CC -1- COFACTOR: FAD AND METAL ION.
 CC -1- PATHWAY: Benzophenanthridine alkaloids biosynthesis.
 CC -1- SUBCELLULAR LOCATION: VESICULAR.
 CC -1- SIMILARITY: Belongs to the oxygen-dependent FAD-linked
 CC oxidoreductase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; S65550; AAB20352.1; -;
 CC DR EMBL; AF005655; AAC39358.1; -;
 CC DR PIR; A41533; A41533.
 CC DR InterPro; IPR001575; Oxid.FAD bind.
 CC DR Pfam; PF01565; FAD binding 4_1.
 CC DR PROSITE; PS00862; OX2_COVAL_FAD; 1.
 CC DR OXidoreductase; Signal; Glycoprotein; Flavoprotein; FAD;
 CC KW Alkaloid metabolism.
 CC FT SIGNAL 1 23
 CC FT CHAIN 24 538
 CC FT BINDING 104 104
 CC FT CARBOHYD 38 38
 CC FT CARBOHYD 423 423
 CC FT CARBOHYD 471 471
 CC FT N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SQ SEQUENCE 538 AA; 59958 MW; 1A505F86A06CDB24 CRC64;

Query Match 9.6%; Score 286.5; DB 1; Length 538;
 Best Local Similarity 24.8%; Pred. No. 1.5e-14;
 Matches 128; Conservative 72; Mismatches 170; Indels 147; Gaps 27;

QY 71 TVRIVSGGCHCEYDFV-DECVKAIINVTGLVSGYD-DIRGVFSSGDTTWMSFKTLFRD 128
 DB 96 TIRLRSGGHSYEGLSYADPPFVILIDMLNLRISIDLESTVAVESGSLTGLIYVAITBS 155
 QY 129 HGRV-LPGSSCYSVGLGHIYVGGDGLIARHGLPVDWISGVEVVPVLT-DSVLKY 185
 DB 156 SSKLGFAGWCPTVGGHISGGGFGFMSKRYGLAAD-----NVDAIILIDANGAIL-- 207
 QY 186 VHMDSGNDGLFWAHTGGGNGFIITKRYFKD.LMSPRGV-----ASNLFH 234
 DB 208 ---DROMGSDVFWALRGGGGVWGAIVAMKIKLIPVEKTVFRVTKVNAIDEATSLH 264
 QY 235 SMDGTRDALQDLTKYFKLACD-----WKNVTG-KFOIFHOAAEEFVM----- 278
 DB 265 KMQ-FVAEBLEBFT-LSVLGADDEKQWLMIGFHFGLKTVAKSFDDLFPGLVEED 322
 QY 279 YLYTSYND-----AREVAODRHYLEADIEQIYVT-CEPYALGAGHAGMAPFVPRPK 332
 DB 323 YLEMSWGESFAYLAGLETWSQLNNRFLKED-ERAFKTKYDLTK-----BFLPSK--- 370
 QY 333 RHRSKTSYMHDETMDPYVALTETTINGSGPNRGKYSKAMIMDPPDFQIDVIMKLTLEV 392
 DB 371 -----AFYGLLER-----LSKE 382
 QY 393 PDGLTAEMDALLQVDFEGEHLKVMVDTAAVQREVI-IKLOYOTYW--OEEDDAVN 449
 DB 383 PNG-----FIALNGFGGMSKISSDTPPHRSCTRIMWEYIVANNQSFQKKTTER 433
 QY 450 LKMIKRDYFEEWYDYGVPDNTQVSGKGVFEGCYFNPVDLN--NMKNY--GAL 505
 DB 434 LDWLEKYEEM-KP-----VSKNPLG-----YNNHIDLDIGIDWGNKTVANNATIE 480
 QY 506 L-----YFLGNLNLRIKAKMLMDPNEIFTNKQSIIP 535
 DB 481 ISRWGESYFLSNYERLIRAKTLIDPNNVFNHPQISIP 517

```

RESULT 3
MCRA_STRLA
ID MCRA_STRLA STANDARD; PRT; 447 AA.
AC P43485;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitomycin radical oxidase (EC 1.5.3.-).
GN MCRA.
OS Streptomyces lavendulae.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=1914;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC STRAIN=B619;
RX MEDLINE=94292477; PubMed=7517396;
RA August P.R., Flickinger M.C., Sherman D.H.;
RT "Cloning and analysis of a locus (mcx) involved in mitomycin C
resistance in Streptomyces lavendulae.";
RL J. Bacteriol. 176:4448-4454(1994).
CC -1- FUNCTION: INVOLVED IN MITOMYCIN RESISTANCE; OXIDIZES REDUCED FORM
OF MITOMYCINS.
CC -1- COFACTOR: FAD (POTENTIAL).
CC -1- INDUCTION: BY MITOMYCIN C IN CONCENTRATIONS AS LOW AS 300 NM.
CC -1- SIMILARITY: Belongs to the oxygen-dependent FAD-linked
oxido-reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L29247; AAA21476.1; -
DR InterPro: IPR001575; Oxid_FAD_bind.
DR Pfam: PF01565; FAD_binding_4; 1.
DR PROSITE: PS00862; OX2_COVAL_FAD; 1.
KW Antibiotic resistance; Oxidoreductase; Flavoprotein; FAD.
FT INIT MET 0
FT BINDING 63 63 FAD (COFACTOR) (BY SIMILARITY).
SQ SEQUENCE 447 AA; 48220 MW; 3B0071B818D838C CRC64;

Query Match 5.2%; Score 154.5; DB 1; Length 447;
Best Local Similarity 22.4%; Pred. No. 0.00018;
Matches 85; Conservative 51; Mismatches 163; Indels 81; Gaps 17;

QY 21 PDKP---DPLPBMKOGFNRWIGTNDIVYVYTPGACTALDRAMEKCSFGTVRVSG 77
DB 10 PDQPGYDARL-----GLNRA-AESRPAYVEADEQEVAALAEQKRP--VGVMAT 61
QY 78 GHCYEDPVECEYKATINVTGLVBSGYDDDRGFVSSGDTNNMGSFKTLFPHDGRVLP 137
DB 62 GH-GPVSADDAV--LVNTRMEGVSYDAAPATAMTEAGRMKRVLEHTAPHGLAPINGS 118
QY 138 CYSVGLGHIIVGGGDLARLHGLPVDMLSGEVNVPVLTEDSVLKYVHKDSENGDEL 197
DB 119 SPNVGAVGVVGGAGLGRFGVADHVRRLV-----TAGGRLRDV---TAGTDPDL 170
QY 198 FMAHTGGGGNGFIIITKYFKDLPMSEPRGVASNLHFSWMDGFTFDALQDILTYFKLARC 257
DB 171 FMA-VGGGKNFGVGVMEVDLPVLT--RLYGGGLYPA-----GEATAEVLYHAAEWVRH 222
QY 258 DWKNTYKGPQIFHQ-----AAEFVWVLYTSYNDAREVAQORHRLHLEADIEQIY 308
DB 223 VPBEAMSSVLLVHNPDLPDVEPLRGFFITHLRIAYSGE-----PADGSHLV 269
QY 309 KTEPTKALGSHGMAFPV---RPRKRTSKTSYMEDETMDEYFVALTETINGSGPNQR 365

```

```

DB 270 R---PLRELG-----PILDDTVADMEYAEVGTIIHETSPMPVYA----- 305
QY 366 GKYSAYMIKDPDPDQIDVI 385
DB 306 --YDRNVLSDLTDAVDII 323

RESULT 4
DIVL_CAUCR
ID DIVL_CAUCR STANDARD; PRT; 769 AA.
AC Q9RQ09; Q9A2S2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sensor protein divl (EC 2.7.3.-).
GN DIVL OR CC3484.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxId=155892;
RN [1]
RP SEQUENCE FROM N.A., AND AUTOPHOSPHORYLATION SITE.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=20027501; PubMed=10557274;
RA Wu J., Ohta N., Zhao J.L., Newton A.;
RT "A novel bacterial tyrosine kinase essential for cell division and
proc. Natl. Acad. Sci. U.S.A. 96:13068-13073(1999).
RL Proc. Natl. Acad. Sci. U.S.A. 96:13068-13073(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
Kolony J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,
Utechtack T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: Required for cell division and growth. It catalyzes the
phosphorylation of CtrA and activates transcription in vitro of
the cell cycle-regulated flp promoter.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: Autophosphorylated.
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF083422; AAP08344.2; -
DR EMBL: AE006007; AAK25446.1; -
DR TIGR: CC3484; -
DR PhosSite; Q9RQ09; -
DR InterPro: IPR004358; Bact_sens_pr_C.
DR InterPro: IPR003661; His_kin.
DR InterPro: IPR004359; His_kin_sis.
DR InterPro: IPR000014; PAS_domain.
DR Pfam: PF02512; HATPase_c; 1.
DR Pfam: PF00518; signal_1.
DR PRINTS: PR00344; BCTRLSENSOR.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HisKA_1.
DR TIGRFAMs: TIGR00229; sensory_box; 2.
DR PROSITE: PS50109; HIS_KIN; 1.
KW Sensory transduction; Transferrase; Kinase; Transmembrane;
Phosphorylation; Complete proteome.

```

FT	TRANSMEM	6	26	POTENTIAL.
FT	DONAIN	547	758	HISTIDINE KINASE.
FT	DONAIN	9	221	ALA-RICH.
FT	MOD_RES	550	550	PHOSPHORYLATION (AUTO-).
FT	CONFLICT	200	200	Q -> H (IN REF. 1).
FT	CONFLICT	216	216	V -> E (IN REF. 1).
SO	SEQUENCE	769 AA;	82796 MM;	002B2428F18A57EF CRC64;
Query Match				
Best Local Similarity 4.0%; Score 118.5; DB 1; Length 769;				
Matches 109; Conservative 69; Mismatches 192; Indels 117; Gaps 30.				
QY	59	ALDRAMKCSGPTGATIVASGHCYEDFVDECKAI---	INTVLGVSQYDROGRFYVSSG	115
DB	58	AFDPAIVIAVEGGRANLVAGG--EGII--ACAKLAGDAEYSAVVA	-----LSDA	103
QY	116	DTNMGSPFTLPRDHRVLPGGSCYSVGLGH---	IVGGDGIILRLHGLPYDML-SGVE	170
DB	104	DPNVAQKLTALFEREE-----PCVPEARGPCHOLVSVEGRAAGALMLRLAPIDRADSLP	158	
QY	171	VVKRFLTEDSVLK--VYHKDSEGDGELFWAH-----	TGGGGNFGIITKYFKDLP	222
DB	159	TAAAPFAFVDSVBEPCWI---AGADGQAIWGNAAFVBAVGAASQAAPALAGKSF-----	209	
QY	223	SPRGVIAENLHFSMNGFTRDALQDILTKYFKLARCQDKNTYCKPQIFFGQAEFF---	VMY	279
DB	210	-DRGADAVAVEAAGKGREREAR-----	WINVERRRARFRLSAQPLDGGGG	255
QY	280	LYTSYSDAREVADRDHYHLEADIEQIYKTC-----	PTVALGSHG--WAPPV	328
DB	256	VFCADVTIE-DYRAFKKHVAHDETILNHAENVAISQIRRLSYHTAPFELMGLBPA	314	
QY	329	RPRKHTSKTSYMHDETD--YFYALTTETINGSQPNRGKYSAYMTKDPPDF--	QIDV	384
DB	315	WLADRP-----HGEVLDRLRQRRLRPETIDYAG-----	WKAEELAR-YEDLGPQAD	361
QY	385	IKWKITETVPDGLTSLAEKMDALLQVMFGEI---	HKVWDATAYVAQREITIKLQYQYQ	441
DB	362	LM---DLPRDRTLKVR---QPRPLGMLIYSDITGIELRLKAQYNALIQVQATTDK	413	
QY	442	EEDKAV-----NLKMRDVEEMEPYGGVDPPTQVBSG--	KGVFEGCYENYPDV-DL	493
DB	414	LNDAAVAVGSQGRLR---LHNEAFETEMNV-TPHALBAADPFEQVELC--	VERLHDJ	465
QY	494	NNVKNKG	500	
DB	466	SFWRELK	472	
RESULT 5				
MM03_RABIT				
ID	MM03_RABIT	STANDARD;	PRT;	478 AA.
AC	P28863;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3)			
DE	(MMP-3) (transin-1) (St-1).			
GN	MMP3.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID=9986;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86077214; PubMed=2825726;			
RA	Finl M.E., Katmilowicz M.J., Rudy P.L., Beeman A.M., Borges K.A.,			
RT	Brinckhoff C.E.,			
RT	"Cloning of a complementary DNA for rabbit proactivator. A			
RT	metalloproteinase that activates synovial cell collagenase, shares			
RT	homology with stromelysin and transin, and is coordinately regulated			
RT	with collagenase."/			
RT	Arthritis Rheum. 30:1254-1264(1987).			

RN	[2]
RP	SEQUENCE OF 1-167 FROM N.A.
RX	MEDLINE=87156645; PubMed=3030290;
RA	Whitham S.E., Murphy G., Angel P., Rahmsdorf H.J., Smith B.,
RA	Llyons A., Harris T.J.R., Reynolds J.V., Herrlich P., Docherty A.J.P.;
RT	"Comparison of human stromelysin and collagenase by cloning and
RT	sequence analysis."
RL	Biochem. J. 240:913-916 (1986).
CC	-I- FUNCTION: CAN DEGRADE FIBRONECTIN, LAMININ, GELATINS OF TYPE I,
CC	III, IV, AND V; COLLAGENS III, IV, X, AND IX, AND CARTILAGE
CC	PROTEOLYCANAS. ACTIVATES PROCOLLAGENASE.
CC	-I- CATALYTIC ACTIVITY: Preferential cleavage where P1', P2' and p3'
CC	are hydrophobic residues.
CC	-I- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M0A.
CC	-I- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcement/)
CC	or send an email to license@ebi.ch).
CC	-----
DR	EMBL; M25664; AAA31467.1; -.
DR	PTR; A37306; KCRBS1.
DR	HSSP; P08254; ISLM.
DR	MEROPS; M10_005; -.
DR	IinterPro; IPRO00585; Hemopexin.
DR	IinterPro; IPRO01818; Matrxin.
DR	IinterPro; IPRO00130; Zn.MTpepdase.
DR	pfam; PF00045; hemopexin; 4.
DR	pfam; PF00413; peptidase_M10; 1.
DR	PRINTS; PR00138; MATRIXIN.
DR	SMART; SM00120; HX; 4.
DR	SMART; SM00235; ZnmC; 1.
DR	PROSITE; PS00024; HEMOPEXIN; 1.
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.
KW	Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW	Collagen degradation; Extracellular matrix; Signal.
FT SIGNAL	1 17 PROBABLE.
FT PROPEP	18 100 ACTIVATION PEPTIDE.
FT CHAIN	101 478 STROMELYSLIN-1.
FT DOMAIN	288 478 HEMOPEXIN-LIKE.
FT SITE	93 93 CYSTEINE SWITCH (POTENTIAL).
FT METAL	219 219 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE	220 220 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL .	223 223 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL .	229 229 ZINC (CATALYTIC) (BY SIMILARITY).
CARBOND	121 121 N-LINKED GLUCNAc. . .) (POTENTIAL).
DISULFD	291 478 BY SIMILARITY.
CONFLICT	83 83 N-> D (IN REF. 2).
CONFLICT	128 128 R-> K (IN REF. 2).
SEQ	478 AA; 53942 MW; CA742E31A4549D40 CRC64;
Query Match	3.9%; Score 116; DB 1; Length 478;
Best Local Similarity	20.7%; Pred. No. 0.19; Indels 182; Gaps 24
Matches 101; Conservative 59; Mismatches 145; Indels 182; Gaps 24	
Db	38 RWIGTNDIFVVVTPQACTALDRAMEKCSPTGTVRIIVSGCHYEDEFDECVKAIINV 97
Oy	: ::: :: :: :
Db	109 KMTHHLTYLRVNTPDLPRAVDAAIEK-----ALKWMEEVT 146
Oy	98 GLVES-GYDDRGYFVSAGDTNWGSFKTLFRDHGRVLPGSGCYVGGLGHIVGGDGILA 156
Db	147 PLTESRKKEGEADIMISFG-----VREHDFIP----- 174
Oy	157 RLHGLPVDWLMSGVEVVKKPVLTEDSVLKYYKHDKSEGNDELFMATHGGGGNFGITTKY 216
Db	175 -----FDGEGN--VLAHAYAPQG----- 191

```

QY 217 FKDLPMSPRGVIAINLHFSWD-GFTFDALQDLLTKYFKLARCDWKNTVSKFOIFHOAAE 275
DB 192 -----INGDAHFDDEBQWTDQTTG---TNLFVAABHLGSLG---LFHSANPE 234
QY 276 FVMY-LYTSYSDNAEVAQDRHYHLEADIEQIYKCEPTKALGAGNAPPPVPRKRKH 334
DB 235 AMYPPVYNFTDLARFLSOD-----YVDGIQSLYGAPASPDRSGVPMEEVPP----- 283
QY 335 TSKTSYMHEDMDYPPVALTERINGSGPNRGKY---KSAVMKDDPPDQ-IDVIRKYL 390
DB 284 GSGTPEMCPD--DLSDAIS-TIRGEILFFKDRIFWRKSLRIIE--PEPHLLSFFMPSLP 338
QY 391 EYVDDGLTSAEMKDALQVDMFGG-----EIKHYVMAT-----AVAQ 427
DB 339 SAVDAAYEYISRTD---VFIFKGTQFMAIRGNVQAGYPSRHTLGPFTIRKIDALISD 395
QY 428 REYIIKQYQTYQWQEDKQAVNLKATRDYEEKYEP-----YGVDPDPTQVSGKGYF 481
DB 396 KE-----RKXTYFVBDK-----YMRFDKROSLBGPRIHIADEPFGINPKIDA---VF 442
QY 482 EGCYENY 488
DB 443 EAFGFY 449

RESULT 6
YE68_METUA STANDARD; PRT; 1009 AA.
ID YE68_METUA
AC 058663;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein MJ1468.
GN MJ1468.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Werrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073 (1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: CONTAINS 5 PKD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67588; AAB99478.1; -.
CC TIGR; MJ1468; -.
DR InterPro: IPR000601; PKD_domain.
DR Pfam: PF00801; PKD; 6.
DR SMART: SM00089; PKD; 7.
DR PROSITE: PS50093; PKD; 5.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 6 26
FT TRANSMEM 985 1005
FT POTENTIAL.
FT TRANSMEM 985 1005
FT POTENTIAL.

```

```

FT DOMAIN 213 247 PKD 1.
FT DOMAIN 436 503 PKD 2.
FT DOMAIN 724 806 PKD 3.
FT DOMAIN 822 886 PKD 4.
FT DOMAIN 925 962 PKD 5.
FT DOMAIN 293 298 POLY-ASN.
SQ SEQUENCE 1009 AA; 115119 MW; 13E9B4933EAB7972 CRC64;

Query Match 3.7%; Score 111; DB 1; Length 1009;
Best Local Similarity 18.3%; Pred. No. 1.2; Indels 222; Gaps 26;
Matches 113; Conservative 70; Mismatches 214;

QY 69 PGNVIVSGGHCY-----EDFVEDECYKALINVTGLVSGYDDRGYF 111
DB 28 PNTNMTNSSGYVNDPNILAVNNDITTFELALPDSVAQDILLD-NGVVKMPFGD----- 81
QY 112 VSSGDTNMGSEFTLFRDHGRVLPQSGCYSGVIGHHIVGGDGLIARLHGLPVDWLSGVEV 171
DB 82 --LLETVDYGNVYRTTHTTYTFFPP-----YFVAMCGYLNNTGVS-----KALTYNWL----- 125
QY 172 VVKPVLTEDSVLKYYHKKSEGN---DGLFMAHNTGGGGNPGIITKYY-----FKDL 220
DB 126 ---VGVDAVNTKTYFVNSGSPLSKTSWEVYINST-----NNTVITIKYSETPVDRENGCL 176
QY 221 PMSPRGVIAINLHFSWDGFTFDAL--QDLITKYFKLAR---CDWKNTVSKFOIFHOAAE 274
DB 177 SVDTTISVTYN-----VSRLDEIVGDTVKNFVSRIIICWMSFGDGTFSF-----E 223
QY 275 EFMVTLYTSYSDNAEVAQDRHYHLEADIEQIYKCEPTKALGAGNAPPPVPRKRKH 334
DB 224 KPEPHYTKSGLYPRVLAVDDSGRVM--VGYLDGIEVKRARGYIYVWTGP----- 274
QY 335 TSKTSYMHEDMDYPPVALTERINGSGPNRGKYKSAVMKDDPPDQIDVIRKYL----- 389
DB 275 ---SHYDGEAVTYVYNSGDDNNNNNAAYTDPYKITYKVNDTIRFEMSGANGERYKMPD 330
QY 390 ---TEVPDGLTSAEMKDALQVDMFGGELHK-----V 418
DB 331 GQGTERTPYTKS-----YTPSHQYKFPFPMWPFPMMSYGMGSMKSDTLNFIY 379
QY 419 VMDATAVAGREYIIKQYQTYQWEE---DKDAVNLKMW-----IDFY----- 457
DB 380 VQDVVENTRYNFPSSAHDKTYDYENKENVNLVYYSVDVISTPKFNKLDGYIDIT 439
QY 458 -----BEMTEPIYGVVD 469
DB 440 ATADKTQYVSNENVRFDSPYGNPIFIWVCFDGTGCSFEKSPTHRYSGLYYPHFVID 499
QY 470 PNTQVSSG--KGVFEGCYENYEDVDLNNMKNKGKYGALBL---YFLGNLRLIKAKYLMWP 524
DB 500 DNGNIEVGIIPPIGVGGYSYFOI-----YASPTIAPTYPIINITIVBPASNTWTW 550
QY 525 NEI-FTNKQSIPTKLPKEP 542
DB 551 HHIVFDGGSVMIKPKKSP 569

RESULT 7
DLD3_YEAST STANDARD; PRT; 496 AA.
ID DLD3_YEAST
AC P39976;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable D-lactate dehydrogenase [cytochrome] (EC 1.1.2.4) (D-lactate
DE fericytochrome C oxidoreductase) (D-ICR).
GN DLD3 OR YEL071W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=6288c / AB9712;
 RA Diereich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.,
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: (R)-lactate + 2 ferriocytochrome c = pyruvate +
 CC 2 ferriocytochrome c.
 CC -1- COFACTOR: FAD (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE FAD-BINDING OXIDOREDUCTASE/TRANSFERASE
 CC FAMILY 4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U018795; AAB65016.1; -.
 DR SGI; S0000797; DLD3.
 DR InterPro; IPR004113; FAD-oxidase_C.
 DR InterPro; IPR001575; Oxid_FAD_bind.
 DR Pfam; PF01565; FAD-binding_4.1.
 DR Pfam; PF02913; FAD-oxidase_C; 1.
 DR Oxidoreductase; Flavoprotein; FAD.
 KW OXIDOREDUCTASE; FLAVOPROTEIN; FAD.
 SQ SEQUENCE 496 AA; 55225 MW; 4809F74EDF07F520 CRC64;
 Query Match 3.7%; Score 109.5; DB 1; Length 496;
 Best Local Similarity 18.0%; Pred. No. 0.62; Indels 227; Gaps 31;
 Matches 110; Conservative 82; Mismatches 191; Indels 227; Gaps 31;
 QY 6 QKDPGYIVD-----VNAGTPDKPPRLPSMKQGNFRWIGTNIIDFVYV 49
 DB 19 KRNPNFKVLDSEDLAVFRSILSNDELINSQAPES---LASFQNDMMKKYRGS---NL 70
 QY 50 VYTPQAGCTALDRAMKESPGTRIV-SGHCTEDF-----VFDEYKAIIVTGLVES 102
 DB 71 ILLP-NSTDKYSKIMKXCNDKLAIVPQGN--TDLVGASVPFDELIVSLRWNNKVRD- 126
 QY 103 GYDDDGQYFVSSGDTVMWGSFKTLFRDHGRVLP-----GGSYGVAGGHIIVGGDGI--- 154
 DB 127 -FPVSGTFCKDAGVWRDAHQFLHDDHLPDLDBRNCCQ---VGVVSTNAGLNL 182
 QY 155 -LARLHGLPYDWLSGVEVYK--PVLTEDSVLKYVHNSGND-GELFWAHTGGGGNGF 210
 DB 183 RYGSLSGN-----VLGEVLV.PNCEIISINIALR---KDNVTGLKQFLI---GAEGTIG 231
 QY 211 IITKYYFKDLPMSPRGVIASNLHFSMDGFRDALQDLITTYFKLARCDMKNTYKQIIFH 270
 DB 232 VVTVGV--δIVAAPKPLNAVFFGIENF-----DTVQKLFWAKSELSILLAEFMD 282
 QY 271 QAABEFPMVLYTYSNDAEREVAQDRHYHLADIEOIKYCEPTKALGHAAGNAPFVRP 330
 DB 283 RGS-----IECTIEVY-----KDL-----PPFLEN 302
 QY 331 RKHSHTKTSYMHDEIMDYPFYALTEITINGSQPNQGRK-----YKSAVMT 374
 DB 303 Q-----HNFYVLIET---SGSNKRHDEKLTAFKLOTTSKLISEGMA 343
 QY 375 KDPDFQIDVIWKYLEVPDGL-----TSAEMKALLQVDMFG-- 412
 DB 344 KDRADF--DRLMTWRKSVPTACNSYGGMYKYSMLQKDLYSVAATERTLNAGGLGDA 401
 QY 413 -----GEIHVVMDATAVNAQREYIKLQYQTYQGEEDKXAVNKMTIDFY 457
 DB 402 PKRVVNSCGYGVHGDGNIH-----LNTAVREF-----TKQIEDLL 436
 QY 458 EEMVEPYGVDPDNTQVESGKGVFEGCYFNVPDVLNNWNGKXGALELYFLNLRLLIK 517

DB 437 EFVVEYIASKKGSISAEHGIGFHKKGLHYRSD-----IEIRFKDI----- 480
 QY 518 AKMLDPNEI 527
 DB 481 -KNHIDPNCI 489
 RESULT 8
 YJCS_ECOLI
 ID YJCS_ECOLI STANDARD; PRT; 661 AA.
 AC P32717;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein yjcs.
 GN YJCS OR B4083.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxId=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KL2 / MG1655;
 RX MEDLINE=94089392; PubMed=8265357;
 RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
 RA Danielis D.,
 RA "Analysis of the Escherichia coli genome. IV. DNA sequence of the
 RT region from 89.2 to 92.8 minutes."
 RL Nucleic Acids Res. 21:5408-5417(1993).
 RN [2]
 RP REVISION TO 614.
 RC STRAIN=KL2 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SIMILARITY: TO PSEUDOMONAS SP. ALKYL SULFATASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00006; AAC3177.1; ALT INIT.
 DR EMBL; AE000482; AAC7044.1; ALT_INIT.
 DR Ecogene; EGI1955; yjcs.
 DR InterPro; IPR001279; Blactamase-like.
 DR Pfam; PF00753; lactamase_B; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 614 614 H -> D (IN REF. 1).
 SQ SEQUENCE 661 AA; 73150 MW; 64D5037ACFE5E455 CRC64;
 Query Match 3.5%; Score 104; DB 1; Length 661;
 Best Local Similarity 20.9%; Pred. No. 2.4; Indels 162; Gaps 28;
 Matches 117; Conservative 73; Mismatches 208; Indels 162; Gaps 28;
 QY 15 DVNAGTPDKPDRLL--PSMKQGNFRWIGT-----NIDPVY-----VVY 51
 DB 93 DINAAPETVNSPLWQSGINGISGLFKVTDMYQVRGODISNITFVESEKGIIVDPLV 152
 QY 52 TPQAGCTALDRAMKESPGTRIVSGGHCYEDFVPEVCYKAIIN-----VTGLVESG 103
 DB 153 TPFAAKALDLYFQHRPKQRIYAVITYTHSHTHYGS--VKGIISSEADVNSGVQVLAAPG 210
 QY 104 YDDR-GYVSSGDTVMWGSFKTLFRDHGRVLPDGSCTSY--GLGHIIVGGGDIILAR--- 157


```

Db 211 FMEDEISENVLAG--NIMSRRLALY-SYGLLLPHNAQGNVNGVLATLGTGDSIIAPTXT 267
Qy 158 -----LHGLPVDML--SGVEV-----VVKPVLTEPSVLKYHK----- 188
Db 268 IYRTGKMIIDGEPFLMTPGSEAPAEKHFYIPALALCTANATHLLNFTLTGAKT 327
Qy 189 -----DSRGNDGE-LFWAHTGGGGNGFI---ITKY-----YFKDLPMSPR 225
Db 328 RDTSKMTVEYLNELTDMWGNDAEVLFPMTWVPWNGNHINDYICKYDITKIYHD----- 381
Qy 226 GVIASNLHSPGDFPTDALODLLTKYFKLARCWKTKVGFQIFQDAEEFWMLYTSYS 285
Db 382 -----OTLHLANQCYTWMIEIGDMIKLPPALAN--NMAARGYGSVSHNAAVYNLYLGYDG 436
Qy 286 NDAEREVAODRHLEADIEQIKTCCEPTKALGSH-----GMAP----- 325
Db 437 NRP-----NLHYGOVEWKERY-----VVALGSSARVNLAEANKQGYRMSABELLKQ 485
Qy 326 -FPEVRPRKRTSKTSYMHDETMYP-----FYALTETINGSGNORGYKSAVMIKD 376
Db 486 VIANPGDDYAKNLIQANNFEQLYQAESATWRGFY-----LTGAKELREGVHKEFSHTTG 540
Qy 377 FDP-----FQIDVIWKYLTVEPDGLTSAEMKDALLQVDMFGEIHKVWMDATAVAGREYII 432
Db 541 SPDTIRGMSVEMLPDFMAVRLDQAKAG--KNISLNFMSNGMDNLNLTNDSVLNRYK--- 596
Qy 433 KLYQYQY---YQOEBDKAV 448
Db 597 TLQPADASFYISREDLHAV 616

RESULT 9
ELM1_ASPFU STANDARD; PRT; 634 AA.
ID ELM1_ASPFU
AC P46074;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Extracellular elastinolytic metalloproteinase precursor (EC 3.4.24.-).
GN MEK.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 530-542.
RC STRAIN=Isolate 13;
RX MEDLINE=95012603; PubMed=7927676;
RA Strakova T.D., Markaryan A., Kolatukudy P.E.;
RT "Molecular cloning and sequencing of the cDNA and gene for a novel
RT elastinolytic metalloproteinase from Aspergillus fumigatus and its
RT expression in Escherichia coli."
RL Infect. Immun. 62:4208-4218(1994).
RN [2]
RP SEQUENCE OF 246-258.
RX MEDLINE=94245315; PubMed=8188335;
RA Markaryan A., Morozova I., Yu H., Kolatukudy P.E.;
RT "Purification and characterization of an elastinolytic
RT metalloproteinase from Aspergillus fumigatus and immunoelectron
RT microscopic evidence of secretion of this enzyme by the fungus
RT invading the murine lung."
RL Infect. Immun. 62:2149-2157(1994).
CC -|- FUNCTION: CATALYZES THE HYDROLYSIS OF ELASTIN.
CC -|- CORFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M36.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
CC EMBL: L29566; AAB07708.1; -.
CC DR MEROPS; M36.001; -.
CC DR InterPro; IPR001842; Fungalysin.
CC DR InterPro; IPR000130; Zn_MTPeptidase.
CC DR Pfam; PF02128; Peptidase_M36; 1.
CC DR PRINTS; PR00999; FUNGALYSIN.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC KM Hydroxylase; Metalloprotease; zinc; signal; zymogen.
CC FT SIGNAL 1 18
CC FT PROPEP 19 245
CC FT CHAIN 246 634
CC FT METAL 429 429
CC FT ACT SITE 430 430
CC FT METAL 433 433
CC SQ SEQUENCE 634 AA; 68726 MW; B94E274BF67F911 CRC64;
CC -----
CC Query Match 3.4%; Score 103; DB 1; Length 634;
CC Best Local Similarity 22.8%; Pred. No. 2.7;
CC Matches 39; Conservative 27; Mismatches 73; Indels 32; Gaps 7;
CC -----
Qy 312 EPTKALGSHGAPFVRPRKHTSKTSYMHDETMYPFVALTETINGSGNORGYKSA 371
Db 151 DPVTLAKGTNTLTQLPITV-----DSASSSETEKESYVFKVSGTVSDPKP-----KLV 200
Qy 372 YMKDFPDFOIDVIWKYLTVEPDG--LTSAEMKDALLQVDMFGEIHKVWMDATAVAGRE 429
Db 201 YFVKD--DGLTALAMVETIDISNMLTYIDAKS-----GEITHGVV-----D 241
Qy 430 YIKLYQYQY--YQOEBDKAVNLKMRDPEEMEPYGVDPNPTQVSGKG 479
Db 242 YVAEADYQVYANGINDPTEGERTVYKDPDVSASEFTWISDSINTTTSRG 292

RESULT 10
ELM2_ASPFU STANDARD; PRT; 634 AA.
ID ELM2_ASPFU
AC P46075;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Extracellular elastinolytic metalloproteinase precursor (EC 3.4.24.-).
GN MEK.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DELTA18;
RX MEDLINE=95231298; PubMed=7715453;
RA Jaton-Ogay K., Paris S., Huere M., Quadroni M., Falchetto R.,
RA Togni G., Latge J.-P., Monod M.;
RT "Cloning and disruption of the gene encoding an extracellular
RT metalloprotease of Aspergillus fumigatus."
RL Mol. Microbiol. 14:917-928(1994).
RN [2]
RP REVISIONS.
RC STRAIN=DELTA18;
RA Sanglard D.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -|- CORFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M36.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

DR EMBL; Z30424; CAA83015.1; -
 DR MEROPS; M36.001; -
 DR InterPro; IPR001842; Fungalysin.
 DR InterPro; IPR000130; Zn_Mtpetidase.
 DR Pfam; PF02128; Peptidase_M36; 1.
 DR PRINTS; PRO0999; FUNGALYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR HydroLase; Metalloprotease; Zinc; Signal; Zymogen.
 KM SIGNAL 1 18
 FT PROPEP 19 245
 FT CHAIN 246 634
 FT METAL 429 429
 FT ACT_SITE 430 430
 FT METAL 433 433
 SQ SEQUENCE 634 AA; 68718 MW; 7BA5A9B70DE184BE CRC64;
 Query Match 3.4%; Score 102; DB 1; Length 634;
 Best Local Similarity 24.0%; Pred. No. 3.3;
 Matches 41; Conservative 26; Mismatches 72; Indels 32; Gaps 8;
 QY 312 EPTKALGGHAGMAPFVRPRKRHTSKTSYMHDETMDFYALTEITINGSGPNRGKXKSA 371
 DB 151 DPTVALKGTNTLTLPITV---DSASSESTEKESEYFKGSGTV--SDP---KAKLV 200
 QY 372 YMKDFPDFOIDVIMKYLTEVPDQ--LTSAMKDALLOVDMFGGEIHKVWDATAVAQRE 429
 DB 201 YFVVD--DGLTALMRVETDIDSNMLTYIDAKS-----GEIHGVV-----D 241
 QY 430 YIKLQIQTY-WQEDKDANLKWIRDFYEMTEPYGVPDPTQVSGKG 479
 DB 242 YVAEADYQVYAMGINDPTEGERTVTKDPWDSVASEFTWISDGTNYTTSRG 292
 RESULT 11
 AMYG CLOSP STANDARD; PRT; 702 AA.
 ID AMYG CLOSP STANDARD; PRT; 702 AA.
 AC P29761;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Glucosylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucanhydrolase).
 GN CGA.
 OS Clostridium sp. (strain G0005).
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1506;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=92339427; PubMed=1633799;
 RA Ohnishi H., Kitamura H., Minowa T., Sakai H., Ohta T.;
 RT "Molecular cloning of a glucosylase gene from a thermophilic
 RT Clostridium and kinetics of the cloned enzyme.";
 RL Eur. J. Biochem. 207:413-418 (1992).
 CC -1- FUNCTION: CGA HAS TYPICAL KINETIC PROPERTIES FOR A GLUCAMYLASE,
 CC BUT THIS BACTERIAL ENZYME HAD HIGHER ISOMALTOSE-HYDROLYZING
 CC ACTIVITY THAN OTHER EUKARYOTIC GLUCAMYLASES.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
 CC -1- MISCELLANEOUS: HIGH ACTIVITY TOWARDS 1,6-GLYCOSIDIC BONDS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

DR EMBL; D12818; BAA02251.1; -
 DR PIR; J00868; J00868.
 DR PIR; S24140; S24140.
 DR InterPro; IPR000165; GH_15.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE; PS00820; GLUCAMYLASE; 1.
 DR HydroLase; Glycosidase; Polysaccharide degradation; Lipoprotein;
 KM Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 702
 FT LIPID 22 22
 FT BINDING 342 342
 FT ACT_SITE 452 452
 FT ACT_SITE 455 455
 FT ACT_SITE 456 456
 SQ SEQUENCE 702 AA; 78658 MW; 955EB4D0AD569546 CRC64;
 Query Match 3.4%; Score 102; DB 1; Length 702;
 Best Local Similarity 20.4%; Pred. No. 3.7;
 Matches 121; Conservative 82; Mismatches 205; Indels 184; Gaps 33;
 QY 21 PDKDPRLLPSMKCGFRRRWIGTINDEYVYVY--TPGACT---ALDRAMEKSPGTVR 73
 DB 43 PGEEDTWASAKQKQ-----VGTANNVSKVWFLANGALSEYVYPTIDADV---ELK 93
 QY 74 -IYSGGHCEYDFVDECVKAIINVTGLVSGYDDRGYFVSGDTWGSF---KTLFRDH 129
 DB 94 FIYTDG--KSFVSDPTKDTI---SKVEKFTDKSLGYKLVNTDKK--GRYRIKFIETD- 144
 QY 130 GRVLPGSGCYSVGLGHHYVGGDGLARLHGLPVMLSGEVYVYKVLTEDSLTKVYHD 189
 DB 145 -----VKNSLTKMKKFA 158
 QY 190 SEG--NDGELEFMAH---TGCGGGNGFIITKYYFKDLPMSRG---VIASNLHFSGMGF 239
 DB 159 LBSGSHDYKLYLYVDHIKQGSYNEGVIYKANNEMLMKAKNDVTALSSNI--GKKGY 216
 QY 240 T-----RDALQDL-----LYKYFKLARCQWNTVQKFOI FQAAEFVWYLYTSYSDN- 287
 DB 217 SIQYVKVNDIMTDLBNKQMTKHYSAR---GNIIIGAEIDLKNSQFELVLSFGSDE 273
 QY 288 ---AEREVADRRHYHLE-ADIEQIYTCETPKALGGHAGMAPP---VRPRKRHTSKTS 339
 DB 274 AVKASIELESENDSKSAVIDEMEKYCNLSNFNKANSLYNSMMILKASDEKTNKGA 333
 QY 340 YMHDETMDFYALTEITINGSGPNRGKXKSAWIMDFPDFOIDVIM-KYLTEVPDGLTS 398
 DB 334 YI--ASLSIFMG-----DQGGDNTGGY-----HLVMSRDLHVANAFTA 371
 QY 399 A-EMKDALLOVDMFGGEIHKVWDATAVAQREYIKLQIQTYWQ--EDKDANLKN----- 452
 DB 372 AGVDVANSLSLDY---LAKVVDNGMIPONTWISKFPWTGIQLEQADPIILSYRLRR 427
 QY 453 -----IRFFY-----EEMTEPYGVPDPTQVSGKGVEGCFYFYPVD-- 492
 DB 428 YDLVDSLVPRLADFIITKMPKQGEERMEIIGV-SPATAAEVAGLTCAAYIAEOKDYE 486
 QY 493 -----LNNWKN-----GKYGALTELYPLGNLRLIKAKVIMDPNEIFT 529
 DB 487 SAQKYGKADNWQKLDINTLTTEHGLE-----NGQYIRIAGLPDPNADFT 533
 RESULT 12
 DPOM NEUIN STANDARD; PRT; 969 AA.
 ID DPOM NEUIN STANDARD; PRT; 969 AA.
 AC P33538;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Probable DNA polymerase (EC 2.7.7.7).
 OS Neurospora intermedia.
 OG Mitochondrion.
 OG Plasmid kalilo.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5142;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92030509; PubMed=1934129;
 RA Chan B.S., Court D.A., Viterula P.J., Bertrand H.;
 RT "The kallo linear senescence-inducing plasmid of Neurospora is an
 RT inversion and encodes DNA and RNA polymerases.";
 RL Curr. Genet. 20:225-237 (1991).
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + [DNA] (N).
 CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
 CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52106; CAA36327.1; ALT_SEQ.
 CC PIR; S17909; S17909.
 CC InterPro; IPR002064; DNA_pol_B.
 CC InterPro; IPR004868; DNA_pol_B_2.
 CC Pfam; PF03175; DNA_pol_B_2; 1.
 CC PRINTS; PR00106; DNAPOLB.
 CC SMART; SM00486; POLB; 1.
 CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 CC Transferrase; DNA-directed DNA polymerase; DNA replication;
 CC DNA-binding; Plasmid; Mitochondrion.
 CC KW DNA-binding; Plasmid; Mitochondrion.
 CC SEQUENCE 969 AA; 112829 MW; 6048B93814F668F CRC64;
 SQ
 Query Match 3.4%; Score 101.5; DB 1; Length 969;
 Best Local Similarity 20.0%; Pred. No. 6.3; Indels 103; Gaps 16;
 Matches 76; Conservative 50; Mismatches 151;
 QY 213 TRYFKDLPMSPRGVIAINLHFSW-----DGFTRDALODLTKYFKLARCWKNV 263
 DB 38 SKFYSTSTD-SPMGHIVSASKSMVYVYSHKNDPLTRSEFHDELKGFSLVKKQFKE-- 94
 QY 264 GKFQIHOAAEEVMTLYTYSNDAR-----EVAODRIYHLEADI 304
 DB 95 -----EKFFMAKIKFNNDIRISISTVOIGSTDPLEVLRLLEAITSTYMHHTG 144
 QY 305 EOIVKTCEPTK-----ALGG--HAGMAPFVPRPKRHTSKTSYVHDETMVPPFALTEIN 358
 DB 145 SE--AVSEFEYELFSLDGLPKGNILTFPKPTSNISIKTYHKNT-----KRNKNIN 197
 QY 359 GSGPNRGKYK-SAYMIKDFPDFOIDVWKYLTVEVDGLTSAMKDALQ-VD-----MFG 412
 DB 198 LSKKNLPLNFKKNGYITPTMDLS-----QWPNHIFINDKNAVLYNNIIISGVNDMLSPF 254
 QY 413 GEHKKVWATAVAQEEYIIKQYQYQWQEDDAVNLKIRDFYEMTEPYGCVDPNT 472
 DB 255 IYINKKYNIITVLYNNTPIFKIKDEKIMSEDDLSSFKRTTTEHQDKVY----- 303
 QY 473 QVESGKGVF--EGCYFNY-----PDVDLNMKNGKYGALLEYFGNINRLIKAKMLDPNE 526
 DB 304 VENGSMVPESEVAVKTSFKIKITRQDLINFEKPIITLTL----- 343
 QY 527 IFTNKQSIPTKPLKEPKQTK 546
 DB 344 ---ETRSVPIHPIKEGDK 360
 RESULT 13
 MCEL_SFVKA
 ID MCEL_SFVKA STANDARD; PRT; 836 AA.

AC P25950;
 DT 01-MAY-1992 (Rel. 22; Created)
 DT 01-MAY-1992 (Rel. 22; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE mRNA capping enzyme large subunit (includes: polynucleotide 5'-
 DE triphosphatase (EC 3.1.3.33) (mRNA 5'-triphosphatase) (Tase); mRNA
 DE guanylyltransferase (EC 2.7.7.50) (GTP--RNA guanylyltransferase)
 DE (GTase)).
 GN D3R.
 OS Shope fibroma virus (strain Kaesa) (SFV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Leporipoxvirus.
 OX NCBI_TaxID=10272;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91306463; PubMed=1649507;
 RA Upton C., Stuart D., McFadden G.;
 RT "Identification and DNA sequence of the large subunit of the capping
 RT enzyme from Shope fibroma virus.";
 RL Virology 183:773-777 (1991).
 CC -1- FUNCTION: CATALYZES THE FIRST TWO REACTIONS IN THE MRNA CAP
 CC FORMATION PATHWAY.
 CC -1- CATALYTIC ACTIVITY: 5'-phosphopolynucleotide + H(2)O =
 CC polynucleotide + phosphate.
 CC -1- CATALYTIC ACTIVITY: GTP + (5')ppp-mRNA = diphosphate +
 CC G(5')ppp-mRNA.
 CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
 CC -1- SIMILARITY: BELONGS TO THE VIRAL GTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M63902; AAA47224.1; -.
 CC PIR; A40478; OOVZRA.
 CC InterPro; IPR004971; Pox_MCEL.
 CC Pfam; PF03291; Pox_MCEL; 1.
 CC Transferrase; Nucleotidyltransferase; Hydrolyase; mRNA processing;
 CC mRNA capping; Multifunctional enzyme.
 CC FT ACT SITE 256 256 GUANYLYLATION SITE (POTENTIAL).
 CC SEQUENCE 836 AA; 97018 MW; 7375FDB54BAE1730 CRC64;
 SQ
 Query Match 3.3%; Score 97.5; DB 1; Length 836;
 Best Local Similarity 18.3%; Pred. No. 11;
 Matches 86; Conservative 75; Mismatches 169; Indels 141; Gaps 21;
 QY 28 LPSMKGFRKRNIGNIDPVYVYTPQACITLADRAEMKSGYRIVSGHCHEDFVD 87
 DB 365 LPSQPEG-----VLFYSQDQNPQYKIKLDTTTHMNIITRYSSSEVIRG 412
 QY 88 ECVKALINVTGLVESGDDRGVYFSSGDTNMGSEFTLRDGRVLPGSGCYS----- 140
 DB 413 E-----NSTPLEYKFKFSDDKF-----PKDYGTGLMLTDVRYNNIYCIATFNIVED 461
 QY 141 VGLGHIIVGGGDDILARLGLPVDMLSGEV--VVKPVLTEDSVLYKHQSEGDGEL 197
 DB 462 VGKKNVV-----PIKIFSEFSAAGELIKPPI--DKTPKLYLKEYGNGQYQI 506
 QY 198 FVAHNGGGGNG-ITTKYFPDLPMSPRGVIAINLHFSWQDFTDALODLTKYKRLAR 256
 DB 507 VVAHIRDQNIKIGDVLDEDKLSDV-----GQHYANDKRYLNPVSYFTN----- 550
 QY 257 CDKNTVGFQIHOAAEEFVMTLY--TSYNDAREV-----ADRRHYLEADIEQ 306
 DB 551 ---KTRGPIGLISNVVTKLLISVCSKTFLDNSKRRKYLALDFGAGALEKPYFEIIS 607
 QY 307 IYKTCPTALGSHGAWAPFVPRPKRHTSKTSYVHDETMVPPFALTEING---SGP 362
 DB 608 LVATPDKEAIG-----KCIERYNSLNSGI--KSKYKFDYIQETIRSVTVYVSSV 655

QY 363 NO---RGKYSAYWIKDPDPDIDVW---KYLTEVPDLTSAEMKALLQVDMGGE- 414
 DB 656 REVPFFPKF-----DLVDMQFAIHSHPHKATATMNNLTETATGAGVLLTMDGL 708
 QY 415 -----IHK-----VMDATAVAQ--REYIHK 433
 DB 709 LSQLTDKKTIVIHKNLPSSEYNSVEKIHEDQLIVNPPSSMRPMGEYIK 759
 RESULT 14
 ID MM01 BOVIN STANDARD; PRT; 469 AA.
 AC P28053;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Interstitial collagenase precursor (EC 3.4.24.7) (Matrix metalloproteinase-1) (MMP-1) (Fibroblast collagenase).
 GN MMP1 OR CLG.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Periodontium fibroblast;
 RX MEDLINE=95201294; PubMed=7894061;
 RA Tamura M., Shimokawa H., Sasaki S.;
 RT "Primary structure of bovine interstitial collagenase deduced from RT cDNA sequence."
 RL DNA Seq. 5:63-66(1994).
 CC -1- FUNCTION: CLEAVES COLLAGENS OF TYPES I, II, AND III AT ONE SITE IN THE HELICAL DOMAIN. ALSO CLEAVES COLLAGENS OF TYPES VII AND X.
 CC -1- CATALYTIC ACTIVITY: Cleaves preferentially one bond in native collagen. Cleavage of the triple helix of collagen at about three-quarters of the length of the molecule from the N-terminus, at 775-Gly-I-Ile⁷⁷⁶ in the alpha-1(I) chain. Cleaves synthetic substrates and alpha-macroglobulins at bonds where P1' is a hydrophobic residue.
 CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
 CC -1- ENZYME REGULATION: CAN BE ACTIVATED WITHOUT REMOVAL OF THE ACTIVATION PEPTIDE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
 CC -1- SIMILARITY: CONTAINS 1 HEMOPLEXIN-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: X58256; CAA41210.1; -;
 DR PIR: S14654; KCB01.
 DR HSP: P03956; IHFC.
 DR MEROPS: M10.001;-;
 DR InterPro: IPR000585; Hemopexin.
 DR InterPro: IPR001818; Matrxin.
 DR InterPro: IPR000130; Zn_MTPpepase.
 DR Pfam: PF00045; hemopexin; 4.
 DR Pfam: PF00413; peptidase_M10; 1.
 DR PRINTS: PR00138; MATRXIN.
 DR SMART: SM00120; HX; 4.
 DR SMART: SM00235; ZnMc; 1.
 DR PROSITE: PS00024; HEMOPLEXIN; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR PROSITE: PS00546; CYSTEINE_SWITCH; 1.
 DR Hydrolase: Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium; Collagen degradation; Extracellular matrix; Signal.
 KW SIGNAL 1 18 BY SIMILARITY.

FT PROPEP 19 99 ACTIVATION PEPTIDE.
 FT CHAIN 100 469 INTERSTITIAL COLLAGENASE.
 FT DOMAIN 275 469 HEMOPLEXIN-LIKE.
 FT SITE 92 92 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 218 218 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 219 219 BY SIMILARITY.
 FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 278 466 PROBABLE.
 FT CARBOHYD 120 120 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 469 AA; 53354 MW; B4A5504CE24BD75 CRC64;
 Query Match 3.2%; Score 97; DB 1; Length 469;
 Best Local Similarity 19.4%; Pred. No. 5.3;
 Matches 61; Conservative 62; Mismatches 98; Indels 94; Gaps 18;
 QY 180 DSLTKV---HDSBGNDELFWAHGSGGNGFIIKTYKFDLPSPRGVLSNHFPM 236
 DB 158 DIMISFPRDHRNSFDP-----GGLA---HAFQ---PAGIGGDHFD 199
 QY 237 DGETRALODLTLYKELARCDKNTVGKFOIFHQAEEFWVLYTSYENDAREVAQDR 296
 DB 200 DEWWTNFPDY--NLRYVAHERGSLG--LAH--STDIGALMPSYTFSGDVQLSD- 251
 QY 297 HYNLEAD-IEQYKTC-PTKALGSHAGWAPPVPRPKHSTKTSYMHDETMDFPYALT 354
 DB 252 ---DIDGIAIYGPSQNPQPVG-----PQTEVCDSKLT-----FDAIT 288
 QY 355 ETINGSGPNRGKYSAYWIKD--FPDQIDVWIKLYTEVPDLTSAEMKALLQVDMG 412
 DB 289 -TIRGEVMPFKDF--YKRTNLYPEVELNFSVWPQPLNGLO----- 330
 QY 413 GEIHKVMDATAVAQREYIHKIQYTMQEDKDAVNLKMRDPEYMEPYG-----G 466
 DB 331 -----AYEADDEYRFFPKNKYMAVKGDV-----LNGYPRDIIRSGFPRTVKS 376
 QY 467 VDPDPTQVESGKVP 481
 DB 377 IDAAVEEDPTGKTYF 391
 RESULT 15
 ID AGAB VIBS7 STANDARD; PRT; 955 AA.
 AC P48840;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Beta-agarase B (EC 3.2.1.81).
 GN AGAB.
 OS Vibrio sp. (strain J70107).
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OC NCBI_TaxID=47913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94250684; PubMed=8193156;
 RA Sugano Y., Matsumoto T., Noma M.;
 RT "Sequence analysis of the agab gene encoding a new beta-agarase from RT vibrio sp. strain J70107".
 RL Biochim. Biophys. Acta 1218:105-108 (1994).
 CC -1- FUNCTION: HYDROLYSES AGAROSE TO YIELD PREDOMINANTLY NEOCALOTETRAOSE AND NEOAGAROHXAOSE.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-galactosidic linkages in agarose, giving the tetramer as the predominant product.
 CC -1- SIMILARITY: BELONGS TO FAMILY 50 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; D21202; BAA04744.1; -
 KW Hydrolase; Glycosidase; Multigene family.
 SQ SEQUENCE 955 AA; 105976 MW; 6E54555DDA8BD215 CRC64;

Query Match 3.2%; Score 97; DB 1; Length 955;
 Best Local Similarity 20.5%; Pred. No. 14;
 Matches 56; Conservative 46; Mismatches 97; Indels 74; Gaps 16;

QY	104	YDDDRGVFVSSGDTNMGSEFTLFRDHGRVLPGGSCVSVGLGHIYGGDGLARLHGLPV	163
DB	629	YDKKKVAVANG--W-----IFGDHARISTGNDYW-----GPIHD-PP	663
QY	164	D--WLSGVEVWVVPVLTEDSVLKYVHKDSGN-----DSELPWAHTGGGGNFGIITKXY	216
DB	664	DPEFVNSVKAMTKLMT-----VDKNPMMWGVVDNEISWNTKNDANHYGLVYNAL	717
QY	217	FKDLPMSP-RGVIASNLHPS-WDGFTRDALQDILITKFKIARCDWKNVTGKFOIFHOAAE	274
DB	718	SYDMKKSPAKAAFTHEHLEKRYW-----AIEDLNT-----SWGKVASNAEFESFD	763
QY	275	EFVMVLYTSYSDAEREVAQDRHYHLEADIEQIYKT--CEPTKALGH-----AGMAPFPV	328
DB	764	-----HRSRLSKNMKKDYAEMLMLSAKYFTVRALKKVLPNHLVLGAPFADWGV	814
QY	329	RPKRHTSKTSYMHDETMDFPFYALTETINGS	361
DB	815	TP---EIAKGAAPYVDVMSYNLYA--EDLNSKG	842

Search completed: June 10, 2003, 09:59:16
 Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:56:56 ; Search time 22 Seconds
(without alignments)
2385.882 Million cell updates/sec

Title: US-09-998-284-2

Sequence: 1 MATLPQKDGXYIVDVNAGT.....IFTNKOSIPTKPKPKPKRKT 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736.5	24.6	685	2 AG0038	probable oxidoredu
2	368.5	12.3	530	2 F86390	hypothetical prote
3	360.5	12.0	367	2 F69816	reticuline oxidase
4	347.5	11.6	539	2 T10626	reticuline oxidase
5	346	11.5	528	2 T10622	hypothetical prote
6	333.5	11.1	529	2 A86391	hypothetical prote
7	328.5	11.0	535	2 F86390	hypothetical prote
8	324.5	10.8	541	2 G86151	F22M8.11 protein -
9	322	10.7	540	2 T10625	reticuline oxidase
10	302.5	10.1	552	2 H86390	reticuline oxidase
11	300.5	10.0	532	2 T00461	TIK7.21 protein -
12	300	10.0	532	2 T10624	probable berberine
13	298.5	10.0	527	2 G86390	reticuline oxidase
14	296	9.9	536	2 F86251	TIK7.22 protein -
15	290.5	9.7	535	2 T07969	hypothetical prote
16	286.5	9.6	538	2 A41533	probable reticuln
17	278.5	9.3	547	2 G70034	reticuline oxidase
18	278	9.3	533	2 A86433	reticuline oxidase
19	277.5	9.3	536	2 A86432	hypothetical prote
20	277	9.2	527	2 E86432	hypothetical prote
21	272.5	9.1	530	2 T10628	hypothetical prote
22	252	8.4	531	2 F86432	hypothetical prote
23	249.5	8.3	540	2 T00463	hypothetical prote
24	242.5	8.1	466	2 T49756	related to berberi
25	240	8.0	527	2 G86432	hypothetical prote
26	238.5	8.0	461	2 A70687	probable oxidoredu
27	207.5	6.9	479	2 H70847	probable oxidoredu
28	203.5	6.8	479	2 B95997	probable oxidoredu
29	190.5	6.4	805	2 A10176	probable oxidoredu

30	185	6.2	431	2 D86433	hypothetical prote
31	180.5	6.0	446	2 G70510	probable oxidoredu
32	154.5	5.2	448	2 A55519	mcrA protein - Str
33	154	5.1	123	2 G70869	hypothetical prote
34	140.5	4.7	461	2 H69350	glycolate oxidase
35	139.5	4.7	511	2 T48777	6-HYDROXY-D-NICOTI
36	118.5	4.0	769	2 B87681	tyrosine kinase D1
37	116	3.9	478	1 KCRBS1	stromelysin 1 (EC
38	113	3.8	592	2 E70455	sulfur oxidation p
39	111	3.7	1009	2 C64483	hypothetical prote
40	110	3.7	781	2 T36143	probable secreted
41	109.5	3.7	496	2 S50518	hypothetical prote
42	109	3.6	934	2 T42394	potassium channel
43	108	3.6	1599	2 S22737	glucosyltransferase
44	108	3.6	2817	2 B97033	uncharacterized pr
45	106.5	3.6	1050	2 JCT578	endo-1,4-beta-xyla

ALIGNMENTS

RESULT 1

AG0038 probable oxidoreductase YP00308 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C/Accession: AG0038

R/ParKhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.,

deno-Tarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, P.; Dougan, G.; F

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B

Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001, MID:21470413, PMID:11586360

A/Accession: AG0038

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-685 <KUR>

A/Cross-references: GB:AL590842; PIDN:CAC89170.1; PID:gl5978409; GSPDB:GN00175

A/Genetic:

A/Gene: YP00308

Query Match 24.6%; Score 736.5; DB 2; Length 685;
Best Local Similarity 33.6%; Pred. No. 3.2e-48;
Matches 195; Conservatve 77; Mismatches 205; Indels 103; Gaps 20;

QY	13	VIVNAGTPDKPDPRLPSMKQGNRRM---IGTNIDFYVYVYTPQACATLDRAMEKSP	69
DB	3	IIDKNVSTVE-----TLQKGFNLKMPNVEQGAETIYICTPDEVFATNTALAGNR	55
QY	70	GTATRVISGGHCVEDFVDECVKAIINVTGLVE-SGYDDRGYVVS-----	114
DB	56	ITVR--SGGCHYGFVSNKSTERTLSIIDGEMSGLDYDEDKTISLMDANKNTYFPSKL	113
QY	115	-GDTNNGSFKTLFRDGRVLPGGSCYSVGLGHIVGGGGLIARLGLPYDMLGVEVVV	173
DB	114	TGNQNNNGVSLYKRGRTIPGSCVSVGGHISGGVGLSLRNLGLTYDWTGVDILV	173
QY	174	KPVLTEDSVLKVKHKSBN-DGELFWAHTGGGNGFIITTKYFKDLPMSPGVIASNL	232
DB	174	PVGNARHLAFRHVRAVSVEVDRELMAACGAGGNFGIILVAYFFDLPVAPOKAWIPL	233
QY	233	HFSMDGFTDALQDILLTKYFKLARCDW-----KNTVKGPFIF-----H-QAAE	274
DB	234	TYFW-----SLKATPPAFK-AYWQMFADNVNATSTEGVNGGLFTILKLNHIDASD	287
QY	275	EFVWVLYTSYSDAREEVAQDRHYHLEADIEQYTKCEPTKALGAGWAP-----FP	327
DB	288	NVLAIQYGPNG---QVGANDIPLNDFIEK-----MNAAGWPTTYDDFILP	334
QY	328	VRRKQHT---SKTSTMDHETMDYPPYALTETINSGPNORGRKKAAYMKDPPQIDV	384
DB	335	NIPFKHLYPGRKIGRTVESASMDWLVHTQMINSGSGSNORGRKYSDDYIKQPSD---EM	391

Qy 385 IWKLTVEPDLITSAEMKDALLOVDMFGGEIHKVWMDATAVAQREYIIKLOQYQOEED 444
 Db 392 CHALLTHTLTATADKRFNQSIVQIDSYGAINSGIATVQNSLKKQYQYWNEN 451
 Qy 445 KDAVNLKMRDYEEMTEPPYGVDPDNTQVESKGVEGCFYFPVDLN-----N 495
 Db 452 DDQTHLTWIRNRYAAV---FHNGKPAPE-----PEGCIYNPDIIDMKYDTSGEEDPN 500
 Qy 496 WKNGYGALELYFLGNLRILKAKMLDPNPFNTKOSIP 535
 Db 501 WNLNYG---WDQOLIKRLIALKARIDPNPIFHELSIP 536

RESULT 2

hypothetical protein TIK.23 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C/Accession: F66390
 R/Theologian, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzalli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11330712
 A/Accession: F66390
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-530 <STO>
 A/Cross-references: GB:AE005172; NID:99797759; PIDN:AAF98577.1; GSPDB:GN00141
 C/Genetics:
 A/Map position: 1
 C/Superfamily: poppy reticuline oxidase

Query Match 12.3%; Score 368.5; DB 2; Length 530;
 Best Local Similarity 27.0%; Pred. No. 3.1e-20;
 Matches 134; Conservative 76; Mismatches 180; Indels 107; Gaps 23;
 Qy 72 VRIYSGGHCYEDPFDEECVKAII-----NVTLVSGYDGRGVSGDTNMGSKFKT 124
 Db 103 LRISSGHDYRGSLYMSVFPVILDMYKRLSIYDVSS-----KKAMIQAGAT-LGELYT 156
 Qy 125 LFRDHGRVL--FGSSCYSVLGHI VGGDGI LARLHGLPVDMLSGVEVVVKPVLTEDSV 182
 Db 157 NVNDVSGQTLAPPAVCATVAGAGHISGGYGNLMRKGIITVDHIDAQII-----DVN 209
 Qy 183 LKYYHNSQEGNDGLFPAHNGGGGNGGIIIT--KYFPDLFMSPRGYASLHFSMNGFT 240
 Db 210 GKLNLRATMGSD--LFMAIRGGGGSGFVILSWKINLVDPKIVT--VFKVXKLTLEGG-- 264
 Qy 241 RDLADLLTKYFKLARCDMKNTVGKF--QIFHOAAE-----FVWYLYTSYND 287
 Db 265 -----TDVLYK-----WQIVASKFPESLFRAMPVQANSTKRGERTIYVFAQELGR 312
 Qy 288 AEREVAODRHYHLEADIEQIYKTCPTKALGHAQMAFPVPRPKRHTSKTSYMHDETM 347
 Db 313 TDALMAIINQWPELGK--HEDCQEMSLNSTLFWMDYPA-----GTPTSIL-----LD 360
 Qy 348 YFPALLETINGSGPNQRGKYSAYMKDPDPQIDIVYWKLTLEVPDGLTSAEMKDALLO 407
 Db 361 RP-----SSPQDFPKSKSDYVKKPIKEGIEKLMKTMLEKNNI-----VWMO 403
 Qy 408 VDMFGGEIHKVWMDATAVAQREYIIKLOQYQOEEDKAVNLKMRDYEEMTEPPYGV 466
 Db 404 FNPYGGVMDRIPLATITAPPHKGNLPIKIQFTTFNANATMSSLSOMKELY-EVAEPPY-- 460
 Qy 467 VPDNTQVESKGVEGCFYFPVDLNMMNKG-----YGALBYFLGNLRILK 518

Db 461 -----VSSNP---REAFNRYRIDVGSNPGSETNVADEAKIYGS--KYFLGNLKRIMDV 508
 Qy 519 KMLMDPNEIFTNKOSIP 535
 Db 509 KAKIDPDPNFFKNQOSIP 525

RESULT 3

reticuline oxidase homolog ygaK - Bacillus subtilis
 C/Species: Bacillus subtilis
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C/Accession: F69816
 R/Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteck
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuno, V.; Carter, N.M.; Choi
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
 Y.M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seior,
 akenuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A/Authors: Yoshikawa, H.F.; Zamestet, E.; Yoshikawa, H.; Darchin, A.
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A/Reference number: A69580; MUID:98044033; PMID:9384377
 A/Accession: F69816
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-367 <KUN>
 A/Cross-references: GB:Z99108; GB:AL009126; NID:G2633055; PIDN:CAB12708.1; PID:el182869;
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: ygaK

Query Match 12.0%; Score 360.5; DB 2; Length 367;
 Best Local Similarity 26.0%; Pred. No. 7.5e-20;
 Matches 118; Conservative 61; Mismatches 157; Indels 117; Gaps 15;
 Qy 106 DGRGVSSGDTNMGSKFKTLFRDHGRVLPFGSSCYSVLGHI VGGDGI LARLHGLPVDW 165
 Db 9 DKLATIYBAGBLEGEYRTLMQ-YGLTPAGTIANGLGLTGIGGIGLTRAAGLTCD 67
 Qy 166 LSGVEVVVKPVLTEDSVLKYNHDSG-----NDGELFPAHNGGGGNGGIIITKY 216
 Db 68 LVQLEMIYA-----DEKEGADLLITVCSNHPDLFWASGGGGGNGGIVTSM 114
 Qy 217 FKDLPMSPRGVIANSLHFSWDEGFTRDALDILTKYFKLARCDMKNTVGKFIQFOAAEF 276
 Db 115 FKAVPISQVSI--STWGMDDFE-----YNTWQ----- 144
 Qy 277 VMYLYTSYNDAREVAODRHYHLEADIEQIYKTCPTKALGHAQMAFP-----VR 329
 Db 145 -----WPEYTD-----RLTSSIEFWPQVRIEALQGFV-----PKTELKLLK 185
 Qy 330 PRKHTSKTSYMHDETMIDYFPYALTTETINGSNGOGKTK--SAYWIKPPDQIDVYK 387
 Db 186 PLKAGSPISGMWKT--PFIEAVTFEFPSPGNOQPOKMRSSFTLEKPLSEPAISTIK 242
 Qy 388 YLTVPEPDLTSAEMKDALLOVDMFGGEIHKVWMDATAVAQREYIIKLOQYQOEEDKDA 447
 Db 243 FLEHAPV-----QNASVWQALGAGAVADQTAFFYRDILAOBYITNWTSGEKR 295
 Qy 448 VNLKMRDYEEMTEPPYGVDPDNTQVESKGVEGCFYFPVDLNMMNKGKYGALBY 507
 Db 296 QNVRWIEGLRTLSKE-----TWMDYVNMWPIEIRNMPRTYGE----- 344
 Qy 508 FLGNLRILKAKMLDPNPFNTKOSIPYKEL 540
 Db 335 ---NVERLRKVTYDYPENVFRFQOSIP--PLR 362

C:Superfamily: poppy reticuline oxidase

Query Match 11.1%; Score 333.5; DB 2; Length 529;
 Best Local Similarity 25.8%; Pred. No. 1.5e-17;
 Matches 128; Conservative 77; Mismatches 184; Indels 107; Gaps 22;

QY 72 VRIVSGHCEYDFVDECVAII-----NVTGLVSGYDDRGYFVSSGDTMGSGFKT 124
 DB 103 IIRISGHDNEGSLYSVSPFVILDMHKLRLITVDSS-----KKAWVQAGAT-LGELYV 156
 QY 125 LFRDHGRVL--PGSGCYSVGLGHIYGGDGILARLHGLPVDMLSGVEVYVVKPVLTEDSV 182
 DB 157 KINASQTLAFPAVCPTVGVGHISSGGYGNLMRKKGITVDHVSQDLI-----DVN 209
 QY 183 LKYVHKGSEGNDELFWAHTGGGNGFIITKYFQDLPMSPR--GVIASNLHFSMDGFT 240
 DB 210 GKILNRSMGED--LFWAIRGGGASFGVILSMKI-NLVKVPKILTVFKYVKLTLEGG-- 264
 QY 241 RDALODLLTKYFKLARCCKNTVGK--QIFHOAAEFV-----MYLYTSYND 287
 DB 265 ----TDVLVK-----MQLVATKFPEDLFMRAMPQILINGERDRTIAVFAQFLGP 312
 QY 288 AERVAADRRHNLADIEQIYKTCPTKALGAGHAGAPFVRPRKRTSKTSYHDETM 347
 DB 313 ADKLLATMORLPFLGLRR--EDCHEMSWFTTTLFMADYPAGTPK-----SVLLDRPTN 364
 QY 348 YPFALYLETINGSPNORGKYSAYMKDPDFQIDVIMKYLTEVPDGLTSAEMKDAL-L 406
 DB 365 PGFF-----KSKSDYVKKPIPKEGLEKIMK-----TWFKNNIVMM 400
 QY 407 QVDMFGEIHKVWDATAVAQRE-YIIKLQYQYMOEDKDAVNLKMRIFYEMEYEPYG 465
 DB 401 QFNRYGGVMDQIPSTATAFPHRKGNMFYQYSTTWLANATEISLMMKELY-KVAPPY- 458
 QY 466 GVDPDPTQVBSGKGVBECCYFNYPVDLNNKNGKYGALH-----YFGLNRLIKAK 519
 DB 459 -----VSSNP--REAFENYRDIDIGSNPSDETVDVDAKIYGYKFFLGNLRLMQVK 507

QY 520 WLMDPNEIFTKOSIP 535
 DB 508 AKYDPENFFKNEQSI 523

RESULT 7

hypochemical protein TIK.24 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C/Accession: E86390
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marzalli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: E86390
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-535 <STO>
 A/Cross-references: GB:AE005172; NID:99797760; PIDN:AAF98578.1; GSPDB:GN00141
 C/Genetics:
 A:Map position: 1
 C:Superfamily: poppy reticuline oxidase

Query Match 11.0%; Score 328.5; DB 2; Length 535;
 Best Local Similarity 27.5%; Pred. No. 3.6e-17;
 Matches 138; Conservative 66; Mismatches 180; Indels 117; Gaps 25;

QY 72 VRIVSGHCEYDFVDECVAII-----NVTGLVSGYDDRGYFVSSGDTMGSGFKT 124
 DB 103 IIRISGHDNEGSLYSVSPFVILDMHKLRLITVDSS-----KKAWVQAGAT-LGELYV 156
 QY 125 LFRDHGRVL--PGSGCYSVGLGHIYGGDGILARLHGLPVDMLSGVEVYVVKPVLTEDSV 182
 DB 157 KIDBASQTLAFPAVGIATVAGAGHISGGYGNLMRKFGTTVDHVIDAEIV-----DVN 209
 QY 183 LKYVHKGSEGNDELFWAHTGGGNGFIITKYFQDLPMSPR--GVIASNLHFSMDGFT 240
 DB 210 GKILNRSMGED--LFWAIRGGGASFGVILSMKI-NLVKVPKILTVFKYVKLTLEGG-- 264
 QY 241 RDALODLLTKYFKLARCCKNTVGK--QIFHOAAEFV-----MYLYTSYND 293
 DB 265 ----TDVLVK-----MQLVANKFPEDLFMRAMPQILINGERDRTIAVFAQFLGP 307
 QY 294 --ODRHYHLEADIEQIY-----KTCPTKALGAGHAGAPFVRPRKRTSKTSYHDE 344
 DB 308 QFGRYDELMETINQSFPELGLRREDQEMSWLNTTLFWAMLP-----GTPKTVLLGR 362
 QY 345 TMDYPFYALYLETINGSPNORGKYSAYMKDPDFQIDVIMKYLTEVPDGLTSAEMKDA 404
 DB 363 T-DVPEF-----KSKSDYVKKPIPKEGLEKIMK-----TWFKNNI 397
 QY 405 L-QVDMFGEIHKVWDATAVAQRE-YIIKLQYQYMOEDKDAVNLKMRIFYEMEY 462
 DB 398 VMLHFNYPGGMDRIPSNTATFPHRKGNLFYQYSTTWLANATEISLMMKELY-EVAP 456
 QY 463 PYGVDPDPTQVBSGKGVBECCYFNYPVDLNNKNGK-----YGALELYFLGNLR 514
 DB 457 PY-----VSSNP--REAFENYRDIDIGSNPSDETVDVDAKIYG--YKFFLGNLKR 502

QY 515 LIRAKWLMDEPNEIFTKOSIP 535
 DB 503 LMDVKAQSDPENFFKNEQSI 523

RESULT 8

G86151
 P22M8.11 protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
 C/Accession: G86151
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marzalli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A/Reference number: A86141; MUID:21016719; PMID:11130712
 A/Accession: G86151
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-541 <STO>
 A/Cross-references: GB:AE005172; NID:98570449; PIDN:AAF76476.1; GSPDB:GN00141
 C/Genetics:
 A:Map position: 1
 C:Superfamily: poppy reticuline oxidase

Query Match 10.8%; Score 324.5; DB 2; Length 541;
 Best Local Similarity 25.4%; Pred. No. 7.3e-17;
 Matches 132; Conservative 73; Mismatches 169; Indels 145; Gaps 26;

QY 76 SGGHCEYD-----FVDECVAIIINVTGLVSGYDDRGYFVSSGDT-----NW 119
 DB 110 SGGHYDYGYSINRFFVLD--MSYLRNIT--VDMS-BDGSAGAWGAGATTGVEVYVNTW 164
 QY 120 GSFRTLFRDHG-RVDPGSCYSYVGLGHIYGGDGILARLHGLPVDMLSGVEVYVVKPVLT 178

Db 165 QSSKT---HGTHGPAGVCPVAGAGHISGGGYGNMIRKYLSDVYVTDKIV----- 214
 Qy 179 EBSVLKYVAKDSEGNDELFWAHTGGGGNGFIITKYEFKDLPMSPRGV----- 228
 Db 215 -DVNGHILDRKSGED--LFWAIGGGGASFGVILSFKILVPPVTVFRVEKLVEN 271
 Qy 229 ASNLHFSW-----DGFTRDALODLTKYFKLARCWMKNTGKFOIFHQAAEFVWY 279
 Db 272 ALDMVHKMGVFAKTSBPLFMRLMLQPV-----RNTT-----QVIRASVVA 313
 Qy 280 LITSYNDAREBAQDRHHLLEADIEQIYKTCPTALGSHAGW---PFPVPRP---K 332
 Db 314 LFLGKSDMSLITLTK---FPELGLKP---ENCTEMTWIOSVMMANNNDATVIKPEILLD 368
 Qy 333 RHSTKSYMHDETMVDPYALTTETINGSQPNQKGYKSAVMIKDPDFQDVIWKLTLEY 392
 Db 369 RNDSDASFL-----KRSDDVEKEISKDGLDFCKLMEAA 403
 Qy 393 PD-GLTSAEMKDALLOVMEGGEIHKVWDATAVAQREYIILQYOTYMOEDKDAVN-- 449
 Db 404 GKLGIV-----FNPYGKMSVATATATPFRHKRLFKYOHMMNKMDPGTDVSSSF 453
 Qy 450 LKMRIDPYEEMTEPYGVDPDNTQVBSGKGVPEGCTFNPDPVL-----NNMKNGK-Y 501
 Db 454 MEKTRSFYSYM-APFTKKNPRHT-----YLNTRDLDIGINSHGPNYSREAEVY 500
 Qy 502 GALELFLGNLRILKAKWLMNDNEIPTNKOSIPTKPLK 540
 Db 501 G-RKTFGENFRLVKYKTAVDENPFRRDQSIPTLPTK 537

RESULT 9

T10625

reticuline oxidase homolog F21C20.180 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Jun-2001

C/Accession: T10625

R/Bevan, M.; Pohl, T.; Weizengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke

submitted to the Protein Sequence Database, June 1999

A/Reference number: Z16991

A/Accession: T10625

A/Molecule type: DNA

A/Residues: 1-540 <BEV>

A/Cross-references: EMBL:AL080254; GSPDB:GN0062; ATSP:F21C20.180

A/Experimental source: cultivar Columbia; BAC clone F21C20

C/Genetics:

A/Map position: 4

C/Superfamily: poppy reticuline oxidase

Query Match 10.7%; Score 322; DB 2; Length 540;
 Best Local Similarity 24.4%; Pred. No. 1,1e-16;
 Matches 125; Conservative 79; Mismatches 172; Indels 136; Gaps 22;
 Qy 72 VRIVSGHCYCE--DFVDECVKAIINVTGLVSGYDD--DGRGVVSSGDTWMSFKTLF-- 126
 Db 113 LKIRSGCHDYDGLSYSDKPF--FILDMNSIRDVSDIASAMISAGATIGEVYIYIWEK 171
 Qy 127 -RDHGVLPGGSCYVGLGSHVGGDGLIARLHGLPVDMLSGEVVVKFVLTEDSVLKY 185
 Db 172 SRVHG--PFAVGCPPTVGVGGHLSGGGYGNMVRKFLGSLVDYVEDAKIV-----DVNGRV 222
 Qy 186 VHKDSGNDGELFWAHTGGGGNGFIITKYEFKDLPMSPRGV-----ASNLHFS 235
 Db 223 LDRKAGED--LFWATGGGSSYGVAVLGYKALVPPSVVTVFRVEQYWDGSAVDMVHK 280
 Qy 236 WDS-----FTRDALODLTKYFKLARCWMKNTGKFOIFHQAAEFVWYLYTSYN 286
 Db 281 MGVSGKTPBNLFWRLHILQVYTRKKYKTVR-----ASVALFLGRADVVALISKEF-- 332
 Qy 287 DAEREVAQDRHHLLEADIEQIYKTCPTKALGSHAGWAPPVPRKRHTSKTSYM-- 342
 Db 333 -----PELGL-----KENCSEMTWFGSALW 353

Qy 343 -----DETMDYPPALTTETINGSQPNQKGYKSAVMIKDPDFQDVIWKLTLEVP-DL 396
 Db 354 MDRNLNATQVDEPVFLDRILDTSS--SFGRKSDVATATPKKGISSLKMKILEKILGL 410
 Qy 397 TSAEMKDALLOVMEGGEIHKVWDATAVAQREYIILQYOTYMOE-----DKDAVNLK 452
 Db 411 V-----FNPYGKMAEVAVNNAKPPFRHKLFKIQYSVVMKENSASIEKGYLNOAK 460
 Qy 453 IRDFYEMTEPYGVDPDNTQVBSGKGVPEGCTFNPDPVL-----NNMKNGK-YGAL 504
 Db 461 V-----LYSFTMGFVSKNPR-----SSYFNFRVDVIGVNDGANSYKEGVEYG-- 503
 Qy 505 ELYFLGNLRILKAKWLMNDNEIPTNKOSIPT 536
 Db 504 RKTFGENFRLVKYKTAVDGPNFRRDQSIPT 535

RESULT 10

H86390

T1K7.21 protein - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C/Accession: H86390

R/Theologis, A.; Ecker, J.R.; Palm, C.U.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

aneen, N.F.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: H86390

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-552 <STO>

A/Cross-references: GB:AE005172; NID:9797757; PIDN:AAF96575.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

C/Superfamily: poppy reticuline oxidase

Query Match 10.1%; Score 302.5; DB 2; Length 552;
 Best Local Similarity 25.3%; Pred. No. 3.6e-15;
 Matches 128; Conservative 72; Mismatches 190; Indels 115; Gaps 24;
 Qy 72 VRIVSGHCYCEPDPFECYK-AIINVTGLVSGYDDDR--GYFVSSGDTWMSFKTLFROH 129
 Db 127 LKIRSGCHDNHGFYSVSPFVILDMNLRSDVNLSRKNANVOAGAT--LGELYKINER 185
 Qy 130 GRVL--PGGSCYVGLGSHVGGDGLIARLHGLPVDMLSGEVVVKFVLTEDSVLKY 187
 Db 186 SGLIAPPAVCPVAGAGHISGGGFENLMRKGITVDHYIDAQII-----DVNGKLN 238
 Qy 188 KDSGNDGELFWAHTGGGGNGFIITKY-----YFRDLPMSPRGVYASNLHFS 235
 Db 239 RAAMGED--LFWA--IRGGSSFGVILSMKINLVEVFKILTVFVNKTLGQ--GTDLKY 293
 Qy 236 WDFETDALODLTKYFKLARCWMKNTV-----GKFOIFHQAAEFVWYLYTSYNDAR 290
 Db 294 WQ-LVANKLPDSI--FTRA--WPTTVGPRGEGTV-----AVFYAQFLGPTDKLM 340
 Qy 291 EVAQDRHHLLEADIEQIYKTCPTKALGSHAGWAPPVPRKR-----HTSKTSYHDE 344
 Db 341 EIMDQFPELIGRED-----CHEMSWLNTTLFWANYPAGPKSILDRPPTSNSV-- 392
 Qy 345 TMDYPPALTTETINGSQPNQKGYKSAVMIKDPDFQDVIWKLTLEVPDGLTSAEMKA 404
 Db 393 -----KSKSDPVKCKIPKKGLEKLTWKTKMFKFNSVS----- 423
 Qy 405 LLQVDMFGEIHKVWDATAVAQRE-YIILQYOTYMOEDKDAVNLKMRIDPYEEMYEP 463

Db 424 -LQFNPGVWMDLIPATATAPPHKGNLFAKQYSTWTFDANATSSILAMNLELF-EVAEP 481
Qy 464 YGVVPDNTQVSGSKGVFECYFNPVDLNMKNKGR-----YGALELYLGNLNL 515
Db 482 Y-----VSSNP---REAFNFRDIDIGSNPSGETNDEAKIYGS--KYLGNLRL 527
Qy 516 IKAKWMDPNEIFTNKOSIPTKPLK 540
Db 528 MDVAKATDPNPFKNQSIIPRYK 552

RESULT 11

T00461
probable berberine bridge enzyme [imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein F1913.2
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 15-Jun-2001
C/Accession: T00461; H84760
R/Runway: S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL data library, April 1998
A/Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.
A/Reference number: Z14160
A/Status: translated from GB/EMBL/DBJ
A/Accession: T00461
A/Molecule type: DNA
A/Residues: 1-532 <ROU>
A/Cross-references: EMBL:AC004238; NID:g3033373; PID:g3033375
A/Experimental source: cultivar Columbia
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.;
Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402: 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: H84760
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-532 <STO>
A/Cross-references: GB:AE002093; NID:g3033375; PID:AA012819.1; GSPDB:GN00139
C/Genetic:
A/Gene: F1913.2; At2g34790
A/Map position: 2
A/Introns: 164/2; 326/3
C/Superfamily: poppy reticuline oxidase

Query Match 10.0%; Score 300.5; DB 2; Length 532;
Best Local Similarity 24.0%; Pred. No. 4.9e-15;
Matches 117; Conservative 88; Mismatches 197; Indels 85; Gaps 22;
Qy 72 VRIISGHCYEDFV---DECVKAIINVTGLVSGYD--DNRGYFVSSGDT--NWGSFKT 124
Db 108 LRLSSGHDVHGSLFVAEDETPPVIVDSKLQVDVDDNSASAAHAGATIGEVYRIQ 167
Qy 125 LFRDHGVLPGSGCYVGLGHIIVGGDGLIARLHGLPVDMLSGEVVVKVLTEDSVLK 184
Db 168 KSOHNG--FPAQGLSSIGIGHLVGAAGVMMKRGAGADNVLDARIV-----DANGQ 218
Qy 185 VYHKDSBNGDELFWAHTGGCGGNGITTKYFQDLPMSPRGVLAASNLHSDGFTDAL 244
Db 219 ILDRPAANGED--VFMAIRGGGSGFVILAWKIKLVPV-PATVTVFTV-----TKTLE 268
Qy 245 QDLLTKVFKLARCWMKNTVGFQIFHOAAEFVWVLYLVSNDAB--REVAODRHVLE 301
Db 269 QDGTKVLTK---WEQIADKLD-----DDLFIYIISPAKSTTKPGNRTISMSYQOFL 318
Qy 302 ADIEQIYKTCPTALGAGHAGMAPFP--VRPRKHTSKTSYVMD--ETMDYFVALTEI- 357
Db 319 GDSRRLLDQVMKS-----FPRLGTLKDCOTEMSWIKSVWYIAGFNSAARL 367
Qy 358 -NSGPNRGKYKSAVMIKDPDFOIDVIMKYLTEVPDGLTSAEMKDALLOV-DMFGGEI 415
Db 368 AGKSLFKNHFAKSDPFVKEPIFVEGLGELMERFILE-----EDSPLTIWNPYGGMM 417

Qy 416 HKVWDATAVAORE-YIIKLOQYQWQE-EDKDAVNLKMRIDFYEMEPYGVGVPDPTQ 473
Db 418 SRSESEIIFPHRNGTLFKIOWLSTWQDGKVBEBRMKRIREMSYMEGYVSKNP----- 472
Qy 474 VESGKVFEGGCFNNPVDVLDNMKNKGYALE---LYFGNLNRLIKAWLMDPNEIFTN 530
Db 473 -----RQAVVNYRDLDLGT-NEGETDARBWGAKYKGNFERLVKIKGEPPDNPFRR 523
Qy 531 KOSIPRK 537
Db 524 EQSVPTK 530

RESULT 12

T10624
reticuline oxidase homolog F21C20.170 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Jun-2001
C/Accession: T10624
R/Bevan, M.; Pohl, T.; Weizengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke,
submitted to the Protein Sequence Database, June 1999
A/Reference number: Z16991
A/Status: translated from GB/EMBL/DBJ
A/Accession: T10624
A/Molecule type: DNA
A/Residues: 1-532 <BEV>
A/Cross-references: EMBL:AL080254; GSPDB:GN00062; ATSP:F21C20.170
A/Experimental source: cultivar Columbia; BAC clone F21C20
C/Genetic:
A/Gene: ATSP:F21C20.170
A/Map position: 4
C/Superfamily: poppy reticuline oxidase

Query Match 10.0%; Score 300; DB 2; Length 532;
Best Local Similarity 23.5%; Pred. No. 5.3e-15;
Matches 129; Conservative 78; Mismatches 196; Indels 146; Gaps 24;

Qy 36 NRRMIGTNIIDFVVVYTPQACATLDRAMEKSPG---TVRIISGHCYE-----DF 84
Db 73 NGRFSAIDVQKVPVILITPVQSDV--QSAVKCARFGIHRTRSGHDVGLSYTHKPF 130
Qy 85 VFPECVKAIINVTGLVSGYD--DNRGYFVSSGDTNMGSEFTLPRDGR-----VLGGSC 138
Db 131 V-----IIDLNLMSITVDVNRVSWVQTAT---IGELYEIGKKNRLLAFPAQVC 179
Qy 139 YSVGLGHIIVGGDGLIARLHGLPVDMLSGEVVVKVLTEDSVLKVYHKDSBNGDEL 198
Db 180 PTGVGGHSGSGGYTLRKHGLADHVDARV-----DARGILERRKGECD--FF 230
Qy 199 WAHTGGGNGFGIITTKYFQDLPMSPRGVLAASNLHSDGFTDALOD-- 246
Db 231 MAIRGGGSGSFVGLSMKIGLINPSTVTVFVNTKSEGLAKIIRWQ-FVADKVSDDL 289
Qy 247 ---LTKYFKLARCWMKNTVGFQIFHOAAEFVWVLYLVSNDABREVAODRHVLE 302
Db 290 FIRWLORYKNMWRASFPG-----LYL-GSVNKLKMNKKEPELGL 332
Qy 303 D-----IEQIKTCPTALGAGHAGMAPFPVRPRKHTSKTSYVMDETMDYFVALTE 355
Db 333 DDCSTMSIISVWPAE--LG-----EPPINVLTK 360
Qy 356 TINGSGPNRGKYKSAVMIKDPDFOIDVIMKYLTEVPDGLTSAEMKDALLOVDMFGGEI 415
Db 361 RTBAS--LAFKASDPVQSPMKTAISKLMRLQ-----PEAHQAQLITPFGGKM 410
Qy 416 HKVWDATAVAORE-YIIKLOQYQWQE-EDKDAVNLKMRIDFYEMEPYGVGVPDPTQ 474
Db 411 SIADYETPPPHRKNYIEIYLNWRGDKYEX--KRWERYVDMSSEFAKSP----- 463
Qy 475 ESGKVFEGGCFNNPVDL-----NMKNKGYALELYFGNLNRLIKAWLMDPNE 526
Db 464 -----RGAVINDRDLGLNIVYKRSKYEBGSKGVK-YFKNPFRVLAVTISVDP 515

QY 527 IFTNKOSIP 535
DB 516 FPCDEQSIP 524

RESULT 13

G86390
T1K7.22 protein - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C/Accession: G86390
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chinn, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: G86390
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-527 <STO>
A/Cross-references: GB:AE005172; NID:g9797758; PIDN:AF98576.1; GSPDB:GN00141
C/Genetics:
A/Map position: 1
C/Superfamily: poppy reticuline oxidase

Query Match 10.0%; Score 298.5; DB 2; Length 527;
Best Local Similarity 25.4%; Pred. No. 6.9e-15;
Matches 138; Conservative 66; Mismatches 200; Indels 137; Gaps 27;

QY 48 YVVTPOGACTALD-----RAMEKSPGT--VRIVSGCHYEDFVDE 88
DB 60 YVYTKRKRSINYOQVAIVAKHSHVQATVCAKANGIOLIRSGHDYEGSLYS 119
QY 89 CYVAII-----NNGVIVSGYDDDRGRVYSSGDTWGSFKLFRDHGVL--PGSGCY 139
DB 120 SVPEVLDMYNLRSTIVDVS-----KKAWVOAGAT-LGELYTKINEASQTLAPPGVCP 173
QY 140 SVGLGHHVIGGGDGIILARLHGLPVDMLSGEVV--VPEVLTEDSVLKVVHDSGNDGELF 198
DB 174 TVGVGGHITGGGNNMRKFGITVDVHVIDAQLIGVNGKLDLRATM-----GED--LF 223
QY 199 MAHTGGGGNGFIITKY-----YFKDLPMSPRGVASNLHFSWDGFTRDALOD 246
DB 224 MAIRGGGASFGVILSMKINLVEPKLITVEFKVSKTLEQG--GDDVLKMQVATKVPED 281
QY 247 L-----LTKYKLARCMDKNTVGEKQIFHOAAEEFVMTLYTSNDAREVAQDRHYN 299
DB 282 LFRAMPQIYKGGKLG--ERTIG--VVF-----FAQFLGPT--DKLMEIMSGSLPE 326
QY 300 LEADIQIYKTCSEPTALGGHAGAPPPVPRKRHTSKTSYMDDEMDVDFYALTING 359
DB 327 LGIRRED-----CHEMSMTTLTFMANYPGVTPTR-----VLLDRP----- 362
QY 360 SGENORGYKYSAYMIKDFPDQIDVIMKYLTEVPDGLTSAMKDALLOQVDFGEIHKV 419
DB 363 STGEFFKSKSDNIKPIKEGLEKIMK-----TMLKFNFWLIEFPYGGVMDRIP 413
QY 420 WQDTAAVQAE-YIKIQYQYNOBEDKDAVNLKIRDFEEMTEPEYGGVDPNTQYESGK 478
DB 414 ATVTAAPIHRRKGNLFNIOYSTIMLDARETKLTMMEKY--EVAGPY-----VSSNP 463
QY 479 GVPEGCVFNYPDVLDN-----NMKNGK--YGAELELYGLNLRILAKMLMDPNEIFTNK 531
DB 464 ----REALFNRDPDIDIGINPSGLNVDEAKIYG--YKFTLGNLKLMDVYKACDPDNFFKNE 518
QY 532 QST 534
|||

DB 519 QST 521

RESULT 14

F86251
hypothetical protein [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
C/Accession: F86251
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chinn, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: F86251
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-536 <STO>
A/Cross-references: GB:AE005172; NID:g4835780; PIDN:AAD30246.1; GSPDB:GN00141
C/Genetics:
A/Map position: 1
C/Superfamily: poppy reticuline oxidase

Query Match 9.9%; Score 296; DB 2; Length 536;
Best Local Similarity 24.2%; Pred. No. 1.1e-14;
Matches 125; Conservative 65; Mismatches 190; Indels 136; Gaps 20;

QY 76 SGGCHCED-----FPDECVKAIINVTGLVSGYDDDRGRVYSSGDTWGSFKTLF 126
DB 110 SGGHDYEGVSYISHVPFVLDMNLRITVDPAESAW-----VGAATLGEVYRITW 162
QY 127 ---RDGRVLPGGSCYVGLGHHVIGGGDGIILARLHGLPVDMLSGEVV--VPEVLTEDSVL 183
DB 163 EKTSHSG--FPACVCTVAGGHHISGGYGNMIRKGLSDVDTDAKID--VNGQVL 216
QY 184 KYVHKDSGNDGELFMAHTGGGGNGFIITKYFYFKDLPMSPRGV-----ASNLH 233
DB 217 ----DRKGGEDEMFALINCGGASFGVILAFKIKLVPPVTVTVREKNLVENATEW 271
QY 234 FSWDG-----FRDLQDLITKYFKLARCMDKNTVGEKQIFHOAAEEFVMTLYTSY 284
DB 272 HKQFVAPKTDPPQLFWRLLIQPTNRKMQTVRAS-----VVALFLGD 313
QY 285 SDAEREVAQDRHYNLEADIHQYKTCSEPTKALGGHAGAPPPVPRKRH--TSKT---S 339
DB 314 QNTWMSMLTKE-----PELGLKKNCTEMTWIOS 343
QY 340 YMEDEMDVDFYALTINGSGPNR--GKYSAYMIKDFPDQIDVIMKYLTEVPD--GL 396
DB 344 VMMANNADATQIKPEILDRNDPATFGKRSDFEKEITKDGDFLFRKMLEVGKIGL 403
QY 397 TSAEMKDALLOQVDFGEIHKVWDATVAQAEEYITKIQYQYNOBEDKDAVN--LKWIR 454
DB 404 V-----FNPYGIMSTVATTTPPHRRKGLYKIHSMNMKDPGTEAEISFLQKAK 453
QY 455 DPEYEMTEPYGVPDNTQYESGKGVFEGCYFNPVVD--NMKNGKYGAELE--LYE 508
DB 454 SFYSYV--APFVTKPNHT-----YINRDLIDGCVNTPGNSNTRVAEVPGRMTF 500
QY 509 LGNINLRILAKMLMDPNEIFTNKOSIPTKPLEPKQ 544
DB 501 GENFRLVAVKTAVDQNFPRBQSIPTLPGRKARR 536

RESULT 15
T07969
probable reticuline oxidase (EC 1.5.3.9) - opium poppy

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:49:11 ; Search time 43 Seconds
(without alignments)
1691.973 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPKDKGVYIVDVNAGT.....IFTNKOSIPFKLPKPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: A.GeneSeq.101002.*
2: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
3: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
4: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
5: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
6: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
7: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2997	100.0	546	22 AAU02192	Synthetic hexose o
2	2997	100.0	546	22 AAB59205	Chondrus crispus he
3	2989	99.7	546	18 AAW20076	Hexose oxidase, an
4	2985	99.6	546	21 AAY83619	Synthetic hexose o
5	635.5	21.2	539	20 AAY39306	Spn protein invol
6	635.5	21.2	539	22 AAB70952	S. spinosa protein
7	368.5	12.3	530	23 ABB91135	Herbicidally activ
8	347.5	11.6	539	23 ABB93022	Herbicidally activ
9	333.5	11.1	529	23 ABB91138	Herbicidally activ
10	328.5	11.0	535	23 ABB91134	Herbicidally activ

11	324.5	10.8	541	23 ABB90805	Herbicidally activ
12	322	10.7	540	23 ABB93021	Herbicidally activ
13	306.5	10.2	545	21 AAY88391	Tetrahymocanabin
14	302.5	10.1	552	23 ABB91137	Herbicidally activ
15	300	10.0	532	23 ABB93020	Herbicidally activ
16	298.5	10.0	527	23 ABB91136	Herbicidally activ
17	296	9.9	536	23 ABB90968	Herbicidally activ
18	294.5	9.8	548	21 AAG41845	Arabidopsis thalia
19	294.5	9.8	541	21 AAG41844	Arabidopsis thalia
20	294.5	9.8	556	21 AAG41843	Arabidopsis thalia
21	292.5	9.8	509	19 AAW55062	Arabidopsis antifu
22	290.5	9.7	535	22 AAB61789	Optum poppy reticu
23	286.5	9.6	538	22 AAB61790	California poppy r
24	278	9.3	526	23 ABB91198	Herbicidally activ
25	277.5	9.3	533	23 ABB91197	Herbicidally activ
26	277	9.2	528	19 AAW55061	Arabidopsis antifu
27	277	9.2	527	23 ABB91194	Herbicidally activ
28	277	9.2	529	22 AAB61787	Sunflower carbohyd
29	275	9.2	529	19 AAW55053	Sunflower leaf ant
30	274	9.1	529	22 AAB61788	Sunflower carbohyd
31	272.5	9.1	525	21 AAG43378	Arabidopsis thalia
32	272.5	9.1	530	21 AAG43377	Arabidopsis thalia
33	272	9.1	529	19 AAW55054	Sunflower leaf ant
34	266	8.9	540	19 AAW55059	Lectuce antifungal
35	265.5	8.9	544	22 AAB81081	Cannabidiolic acid
36	264.5	8.8	421	21 AAG43379	Arabidopsis thalia
37	263	8.8	508	19 AAW55060	Arabidopsis antifu
38	255	8.5	495	20 AAY24919	Microdochium nivale
39	252	8.4	531	23 ABB91195	Herbicidally activ
40	249.5	8.3	474	21 AAG29445	Arabidopsis thalia
41	249.5	8.3	540	21 AAG29444	Arabidopsis thalia
42	249.5	8.3	551	21 AAG29443	Arabidopsis thalia
43	240	8.0	527	21 AAG42979	Herbicidally activ
44	240	8.0	527	23 ABB91196	Herbicidally activ
45	240	8.0	534	23 AAG42978	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AAU02192	standard; Protein; 546 AA.
ID	AAU02192	
XX	AC	AAU02192;
XX	XX	12-SEP-2001 (first entry)
XX	XX	Synthetic hexose oxidase (HOX) amino acid sequence.
XX	XX	HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW	KW	food manufacturing; Beverage; detergent; baking; dough improving agent;
KW	KW	D-hexose:O2-oxidoreductase; EC 1.1.3.5.
XX	XX	Chondrus crispus.
OS	OS	Synthetic.
XX	XX	WO200138544-A1.
XX	XX	31-MAY-2001.
XX	XX	24-NOV-2000; 2000MO-IB01886.
XX	XX	24-NOV-1999; 99GB-0027801.
XX	XX	(DANI-) DANISCO AS.
XX	XX	Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
XX	XX	Zargahi MR;
XX	XX	WPI; 2001-367695/38.
XX	XX	N-PSDB; AAS06173.

PT Releasing soluble or membrane associated intracellular protein from a
PT cell for manufacturing food, comprises contacting the cell with a
PT membrane extracting composition and causing the protein to be released
PT in soluble form -

PS Claim 16; Fig 6; 108bp; English.

XX The sequence represents the amino acid sequence of synthetic hexose
XX oxidase (D-hexose:O₂-oxidoreductase, EC 1.1.3.5), also referred to as HOX.
CC The native HOX gene was altered using site-directed mutagenesis in order
CC to match the codon usage to known codon preferences of biotechnologically
CC relevant yeasts, such as *Pichia sp.*, to facilitate high level production
CC of HOX in these organisms. The invention involves a method for releasing
CC a soluble or membrane associated intracellular protein of interest (POI)
CC from a cell involving contacting a cell comprising a soluble or membrane
CC associated intracellular POI with a membrane extracting composition (I)
CC and causing the POI to be released from the cell in a soluble form. The
CC method is useful for releasing POI, such as an interleukin 1 receptor
CC antagonist (IL-1ra) which involves contacting a transformed cell
CC comprising IL-1ra with (I) and causing IL-1ra to be released from the
CC transformed cell, in a soluble form. The method is also useful for
CC screening mutated cells or transformed cells producing elevated levels of
CC intracellular POI. The method is used to release a POI for manufacturing
CC food products, such as beverages, preparation of detergents, and in
CC baking as a dough improving agent. The method obtains a fast, specific
CC and economically efficient extraction of a soluble or membrane associated
CC intracellular POI without the use of conventional cell disruption
CC techniques. The resulting cell extract contains less contaminating
CC intracellular DNA and is relatively free of cell wall fragments. The
CC intracellular POI can be recovered from a eukaryotic host organism such
CC as yeast, before glycosylation takes place. The method can be used to
CC prevent contact of intracellular POI with the extracellular growth
CC medium.

XX Sequence 546 AA;

Query Match 100.0%; Score 2997; DB 22; Length 546;
Best Local Similarity 100.0%; Pred. No. 3.9e-281;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATLPQKDGXYIVDVAAGTDPKPRPLPSMKOGFNRRMIGTNDIVYVYVTPQGACTAL 60
DB 1 MATLPQKDGXYIVDVAAGTDPKPRPLPSMKOGFNRRMIGTNDIVYVYVTPQGACTAL 60
QY 61 DRAMEKSPGTIVRIVSGGHCYEDFVDECVKAIINVTGLVSGYDDDRGYVSSGDTNMG 120
DB 61 DRAMEKSPGTIVRIVSGGHCYEDFVDECVKAIINVTGLVSGYDDDRGYVSSGDTNMG 120
QY 121 SFKTLFRDHGRVLPQGSVCYVGLGHIYVGGDGIARLHGLPVDMLSGVEVVVKPVLTED 180
DB 121 SFKTLFRDHGRVLPQGSVCYVGLGHIYVGGDGIARLHGLPVDMLSGVEVVVKPVLTED 180
QY 181 SVLKVYHKDSEGNDELFWAHTGGGNGFGIITKTYFDLPMSRGVIASNLHFSWNGFT 240
DB 181 SVLKVYHKDSEGNDELFWAHTGGGNGFGIITKTYFDLPMSRGVIASNLHFSWNGFT 240
QY 241 RDALODLLTKYFKLARCDMKVTGKFOIFHOAAEFVWYLTYSNSNAERVAADRRYHL 300
DB 241 RDALODLLTKYFKLARCDMKVTGKFOIFHOAAEFVWYLTYSNSNAERVAADRRYHL 300
QY 301 EADIEQIYKTCPTKALGAGHAWAPFVPRPKRHTSKTSYMHDETMVPPYALTELINGS 360
DB 301 EADIEQIYKTCPTKALGAGHAWAPFVPRPKRHTSKTSYMHDETMVPPYALTELINGS 360
QY 361 GPNOGKKSAXVIMKDFPDFOIVTWKLTVEPGLISAEMKDALLOYDNFGGSIHKVYW 420
DB 361 GPNOGKKSAXVIMKDFPDFOIVTWKLTVEPGLISAEMKDALLOYDNFGGSIHKVYW 420
QY 421 DATAVOREYIIXOYOTVWOEDKDAVNLMKIRDFEWEVPEYGVYVDDPTOVESGKV 480
DB 421 DATAVOREYIIXOYOTVWOEDKDAVNLMKIRDFEWEVPEYGVYVDDPTOVESGKV 480
QY 481 PEGCYFNPVDVLLNNWKNKGYLELYFLGNLRLIRAKWLMDPNEIFTNKQSIPTKPLK 540

DB 481 PEGCYFNPVDVLLNNWKNKGYLELYFLGNLRLIRAKWLMDPNEIFTNKQSIPTKPLK 540
QY 541 EPKQTK 546
DB 541 EPKQTK 546

RESULT 2

ID AAB59205 standard; Protein; 546 AA.

XX AAB59205;

DT 23-MAR-2001 (first entry)

DE Chondrus crispus hexose oxidase enzyme protein.

KM Hexose oxidase; marine alga; anti-fouling.

OS Chondrus crispus.

PN W020075293-A2.

PD 14-DEC-2000.

PF 02-JUN-2000; 2000MO-IB00829.

PR 04-JUN-1999; 99GB-0013050.

PA (DANI-) DANISCO AS.

PI Poulsen CH, Kragh KM;

DR WPI; 2001-112148/12.

PT New anti-fouling composition, useful as a coating for treating
PT different surfaces, e.g. outdoor woodwork, external surface of a
PT central heating system, or a hull of a marine vessel -

PS Claim 5; Page 35-36; 36pp; English.

XX The present invention relates to a new anti-fouling composition. The
XX composition involves a surface coating material, a hexose oxidase
XX enzyme obtained from a marine organism and a substrate for the
XX enzyme. The anti-fouling composition is useful as a coating formulated
XX for treating a surface, e.g. outdoor wood work, external surface of a
XX central heating system, or a hull of a marine vessel. It is
XX also useful as an anti-fouling agent for marine structures exposed to
XX seawater flora and fauna.

SQ Sequence 546 AA;

Query Match 100.0%; Score 2997; DB 22; Length 546;
Best Local Similarity 100.0%; Pred. No. 3.9e-281;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATLPQKDGXYIVDVAAGTDPKPRPLPSMKOGFNRRMIGTNDIVYVYVTPQGACTAL 60
DB 1 MATLPQKDGXYIVDVAAGTDPKPRPLPSMKOGFNRRMIGTNDIVYVYVTPQGACTAL 60
QY 61 DRAMEKSPGTIVRIVSGGHCYEDFVDECVKAIINVTGLVSGYDDDRGYVSSGDTNMG 120
DB 61 DRAMEKSPGTIVRIVSGGHCYEDFVDECVKAIINVTGLVSGYDDDRGYVSSGDTNMG 120
QY 121 SFKTLFRDHGRVLPQGSVCYVGLGHIYVGGDGIARLHGLPVDMLSGVEVVVKPVLTED 180
DB 121 SFKTLFRDHGRVLPQGSVCYVGLGHIYVGGDGIARLHGLPVDMLSGVEVVVKPVLTED 180
QY 181 SVLKVYHKDSEGNDELFWAHTGGGNGFGIITKTYFKDLPMSPRVIVASNLHFSWNGFT 240
DB 181 SVLKVYHKDSEGNDELFWAHTGGGNGFGIITKTYFKDLPMSPRVIVASNLHFSWNGFT 240

QY 241 RDALODLTKYFKLARCDMKNVTGKFOIFHOAAEEFWMTLYTSYSDAREVAQDRHYH 300
 XX |||||
 Db 241 RDALODLTKYFKLARCDMKNVTGKFOIFHOAAEEFWMTLYTSYSDAREVAQDRHYH 300
 QY 301 EADIEIYKTCPTALGSHAGMAPPPVPRKRHTSKTSYMHDETMDFYFALTEITNGS 360
 XX |||||
 Db 301 EADIEIYKTCPTALGSHAGMAPPPVPRKRHTSKTSYMHDETMDFYFALTEITNGS 360
 QY 361 GPNORGKYSAYWIKDPFPOIDVIMKYLTEVPDGLTSAMKDALLQVMFGEIHKVVM 420
 Db 361 GPNORGKYSAYWIKDPFPOIDVIMKYLTEVPDGLTSAMKDALLQVMFGEIHKVVM 420
 QY 421 DATAVAQREYIIKLOQYQTYWQEEEDKAVNLKMRDPEYEMYPEYGVPPDNTQVESGKV 480
 Db 421 DATAVAQREYIIKLOQYQTYWQEEEDKAVNLKMRDPEYEMYPEYGVPPDNTQVESGKV 480
 QY 481 FECCYFNYPDVLDLNMKNKGKYGALIEYFGINLRILKAKWLPDNEIFTNKOSIPTKPLK 540
 Db 481 FECCYFNYPDVLDLNMKNKGKYGALIEYFGINLRILKAKWLPDNEIFTNKOSIPTKPLK 540
 QY 541 EPKQTK 546
 Db 541 EPKQTK 546
 RESULT 3
 AAY20076
 ID AAY20076 standard, Proteins 546 AA.
 XX
 AC AAY20076
 XX
 DT 12-SEP-1997 (first entry)
 XX
 DE Hexose oxidase, an antimicrobial/antioxidant agent.
 XX
 KM Hexose oxidase; Chondrus crispus; marine algae; recombinant;
 KM antimicrobial; antioxidant; food preparation; dairy product; starch;
 KM beverage; animal feed; silage; sugar reduction; cosmetics; dental;
 KM toothpaste; dough; lactone production.
 XX
 OS Chondrus crispus.
 XX
 PN P0640905-A1.
 XX
 PD 19-DEC-1996
 PF 04-JUN-1996; 96WO-DK00238.
 XX
 PR 07-JUN-1995; 95US-0476910.
 XX
 PA (BIOT-) BIOTEKNOLOGISK INST.
 XX
 PI Hansen OC, Stougaard P;
 XX
 WI MPI; 1997-052332/05.
 DR N-PSDB; AAT76552.
 XX
 PT Recombinant production of polypeptide having hexose oxidase activity
 XX - used in food preparations as antibacterial and antioxidant agent
 PS Example 3.5; Page 95-97; 124pp; English.
 XX
 OS AAY20076 shows the hexose oxidase (HO) enzyme of the marine algae
 CC species Chondrus crispus. HO is useful in the production of food
 CC products, e.g. dairy products, starch-containing food products (dough)
 CC and non-dairy beverages. HO and active peptide fragments have
 CC antimicrobial and antioxidant properties and act by removing all the
 CC oxygen in a food packaging. HO and peptides of HO can also be used in an
 CC animal feed, especially silage. Further uses are to reduce, or analyse,
 CC the sugar content in a food, in the production of cosmetics, tooth care
 CC products or a pharmaceutical product and in lactone production. HO can
 CC be recombinantly produced in industrially appropriate quantities, and at
 CC a quality and purity level which renders the polypeptide suitable for

CC industrial purposes.
 XX
 SQ Sequence 546 AA.
 Query Match 99.7%; Score 2989; DB 18; Length 546;
 Best Local Similarity 99.8%; Pred. No. 2.4e-280;
 Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MATLPKQDGYIVIDVNACTPPKPPRLPSMKGFRNRIGTNDIVVVVYVYVPOGACTPL 60
 Db 1 MATLPKQDGYIVIDVNACTPPKPPRLPSMKGFRNRIGTNDIVVVVYVYVPOGACTPL 60
 QY 61 DRAMEKSPGTVRIYVSGGICEDFVFEDECVAKINVTGLVSGYDDRCGYVSSGPTNG 120
 Db 61 DRAMEKSPGTVRIYVSGGICEDFVFEDECVAKINVTGLVSGYDDRCGYVSSGPTNG 120
 QY 121 SKRTLFRDGRVLPFGSCYSVGLGGHIVGGGCIILARLHGLFVDMLSGVEVYVVKPVLTE 180
 Db 121 SKRTLFRDGRVLPFGSCYSVGLGGHIVGGGCIILARLHGLFVDMLSGVEVYVVKPVLTE 180
 QY 181 SVLKTVYHSDSEGNDELFWAHTGGGGNGFIITKYYFKDLPMSPRGVILASNLHFSMDGFT 240
 Db 181 SVLKTVYHSDSEGNDELFWAHTGGGGNGFIITKYYFKDLPMSPRGVILASNLHFSMDGFT 240
 QY 241 RDALODLTKYFKLARCDMKNVTGKFOIFHOAAEEFWMTLYTSYSDAREVAQDRHYH 300
 Db 241 RDALODLTKYFKLARCDMKNVTGKFOIFHOAAEEFWMTLYTSYSDAREVAQDRHYH 300
 QY 301 EADIEIYKTCPTALGSHAGMAPPPVPRKRHTSKTSYMHDETMDFYFALTEITNGS 360
 Db 301 EADIEIYKTCPTALGSHAGMAPPPVPRKRHTSKTSYMHDETMDFYFALTEITNGS 360
 QY 361 GPNORGKYSAYWIKDPFPOIDVIMKYLTEVPDGLTSAMKDALLQVMFGEIHKVVM 420
 Db 361 GPNORGKYSAYWIKDPFPOIDVIMKYLTEVPDGLTSAMKDALLQVMFGEIHKVVM 420
 QY 421 DATAVAQREYIIKLOQYQTYWQEEEDKAVNLKMRDPEYEMYPEYGVPPDNTQVESGKV 480
 Db 421 DATAVAQREYIIKLOQYQTYWQEEEDKAVNLKMRDPEYEMYPEYGVPPDNTQVESGKV 480
 QY 481 FECCYFNYPDVLDLNMKNKGKYGALIEYFGINLRILKAKWLPDNEIFTNKOSIPTKPLK 540
 Db 481 FECCYFNYPDVLDLNMKNKGKYGALIEYFGINLRILKAKWLPDNEIFTNKOSIPTKPLK 540
 QY 541 EPKQTK 546
 Db 541 EPKQTK 546
 RESULT 4
 AAY83619
 ID AAY83619 standard, Protein; 546 AA.
 XX
 AC AAY83619;
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Synthetic hexose oxidase.
 XX
 KM Hexose oxidase; production; fermentation; synthetic; modification;
 KM prokaryote; eukaryote.
 XX
 OS Synthetic.
 XX
 PN EP1008651-A2.
 XX
 PD 14-JUN-2000.
 XX
 PF 01-DEC-1999; 99EP-0204068.
 XX
 PR 09-DEC-1998; 98DK-0001630.
 XX
 PA (BIOT-) BIOTEKNOLOGISK INST.

XX Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
 PI WPI, 2000-389309/34.
 DR N-PSDB; AA294011.
 XX Nucleic acid fragment useful for producing large amounts of hexose
 PT oxidase comprises nucleotide sequence encoding hexose oxidase that is
 PT modified by at least one codon
 XX
 PS Disclosure; Page 28-29; 42pp; English.
 XX
 CC Wild type hexose oxidase can be produced by isolating a nucleotide
 CC sequence coding for hexose oxidase naturally produced by an organism
 CC and modifying the sequence so that the modified sequence is
 CC expressed, under identical conditions, at a level at least 10% higher
 CC than the non-modified sequence. The method is useful for producing
 CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
 CC of producing hexose oxidase were to isolate the enzyme from a source
 CC that naturally produces the enzyme, using host organisms which
 CC produce a relatively low level of expression. Therefore, industrial
 CC production of the enzyme is not feasible using this method. This is a
 CC much improved method which produces at least 250 mg enzyme per liter
 CC of fermentation medium.
 CC
 SQ Sequence 546 AA;
 Query Match 99.6%; Score 2985; DB 21; Length 546;
 Best Local Similarity 99.5%; Pred. No. 5.7e-280;
 Matches 543; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MATLPQKPGYIVIDVNAAGTDPKDPRLPSMKOGFNRMIGTINIDFVYVYTPGACTAL 60
 DB 1 MATLPQKPGYIVIDVNAAGTDPKDPRLPSMKOGFNRMIGTINIDFVYVYTPGACTAL 60
 QY 61 DRAMEKSPGTVRIYVSGHCYEDFVDECVAKIINVTGLVBSGYDDRGYFVSSGDTNWG 120
 DB 61 DRAMEKSPGTVRIYVSGHCYEDFVDECVAKIINVTGLVBSGYDDRGYFVSSGDTNWG 120
 QY 121 SFKTLFDRHGRVLPVGGSCYVGLGHTVGGDGLIARLHGPVMTLSGVEVVRPVLTED 180
 DB 121 SFKTLFDRHGRVLPVGGSCYVGLGHTVGGDGLIARLHGPVMTLSGVEVVRPVLTED 180
 QY 181 SVLYVHDSNGNGELFMAHTGGGNGFGITTYFEDLPMSRGYASNLHFSWMDGFT 240
 DB 181 SVLYVHDSNGNGELFMAHTGGGNGFGITTYFEDLPMSRGYASNLHFSWMDGFT 240
 QY 241 RDALQDLITTKYFKLARCDWKTGKFOIFHOAAEFVWYLYTYSYSDAEREVAQDRHYHL 300
 DB 241 RDALQDLITTKYFKLARCDWKTGKFOIFHOAAEFVWYLYTYSYSDAEREVAQDRHYHL 300
 QY 301 EADLEQIYKTEPRLKALGGHAGMAPPPVRPRKRTSKTSYVHDETMVYPFALTEITNGS 360
 DB 301 EADLEQIYKTEPRLKALGGHAGMAPPPVRPRKRTSKTSYVHDETMVYPFALTEITNGS 360
 QY 361 GPNORGKYSAYMTKDPFDIVIMKYLEVPPGLTSAEMKDALLOVDMGGEIHKYVW 420
 DB 361 GPNORGKYSAYMTKDPFDIVIMKYLEVPPGLTSAEMKDALLOVDMGGEIHKYVW 420
 QY 421 DATAVQREYIKIQYQTYWQEDBKDAVNLMKIRDFEEMTEPYGVDPTQVSESGKV 480
 DB 421 DATAVQREYIKIQYQTYWQEDBKDAVNLMKIRDFEEMTEPYGVDPTQVSESGKV 480
 QY 481 FEGGVFNPDVLDNMWKKYGALELYFLGNLRLIRAKMLMDNEIFTNKQSIPTPKL 540
 DB 481 FEGGVFNPDVLDNMWKKYGALELYFLGNLRLIRAKMLMDNEIFTNKQSIPTPKL 540
 QY 541 EPKQTK 546
 DB 541 EPKQTK 546
 RESULT 5

AAV39306
 ID AAV39306 standard; Protein; 539 AA.
 XX
 AC AAV39306;
 XX
 DT 01-DEC-1999 (first entry)
 XX
 DE Spn^u protein involved in spinosyn biosynthesis.
 XX
 KW Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
 KW microtubules; arachnid; nematode; insect; polyketide; polyketide synthase;
 KW PKS; extender module; initiator module; acyl transferase domain; AT;
 KW acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
 KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
 KW insecticide; oxido-reduction.
 XX
 OS Saccharopolyspora spinosa.
 XX
 PN WO9946387-A1.
 XX
 PD 16-SEP-1999.
 XX
 PF 16-FEB-1999; 99WO-US03212.
 XX
 PR 09-MAR-1998; 98US-0036987.
 XX
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX
 PI Baltz RH, Broughton MC, Crawford KP, Madduri K, Merlo DJ;
 PT treadaway PU, Turner JR, Waldron C;
 XX
 DR WPI; 1999-551414/46.
 DR N-PSDB; AA221501.
 XX
 PT New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
 PT for production of insecticidal spinosyn compounds
 XX
 PS Claim 1; Page 153-154; 190pp; English.
 XX
 CC This is the amino acid sequence of the product of the spn^u gene. The
 CC protein is involved in spinosyn biosynthesis. The spn^u gene is one of 23
 CC genes and open reading frames contained in an 80kb DNA sequence
 CC AA221501. Spinosyns are insecticidal microtubules which are useful for the
 CC control of arachnids, nematodes and insects. Biosynthesis of spinosyns
 CC occurs via stepwise condensation and modification of carboxylic acid
 CC precursors generating a linear polyketide which is modified further. The
 CC DNA sequence contains a central region of approximately 55kb which has
 CC homology to the DNA encoding the polyketide synthases (PKS) of known
 CC microtubule producers. The spinosyn PKS DNA region consists of 5 ORFs with
 CC stop codons at the end of acyl carrier protein (ACP) domains. Together
 CC the PKS polypeptides (AAV39297-V39301), form a complex consisting of an
 CC initiator module, spn^u, and several extender modules spn^b-spn^f. The
 CC products of the genes present in the region upstream of the PKS genes
 CC have been assigned names spn^f-spn^s AAV39302-V39315 and are responsible
 CC for different modifications in spinosyn biosynthesis. There are also two
 CC ORFs ORF15 and ORF16 present immediately upstream of spn^s, producing
 CC polypeptides AAV39316-V39317, and two ORFs ORF1 and ORF2 present
 CC downstream of the PKS region producing polypeptides AAV39318-V39319. It
 CC is suggested that spn^u is involved in oxido reduction during spinosyn
 CC biosynthesis. The genes are useful to improve yields of spinosyn, and
 CC for creating new spinosyns e.g. by mutagenesis, or interruption of steps
 CC in spinosyn biosynthesis. The modified spinosyns may be a new insect
 CC control agent or serve as substrates for further chemical modification
 CC and the creation of new semi-synthetic spinosyns. The genes are also
 CC useful to isolate similar sequences from S. spinosa or other species by
 CC hybridization.
 XX
 SQ Sequence 539 AA;
 Query Match 21.2%; Score 635.5; DB 20; Length 539;
 Best Local Similarity 30.2%; Pred. No. 2.5e-52;
 Matches 174; Conservative 91; Mismatches 216; Indels 95; Gaps 16;

QY 2 AILPQKPGYIVDVNAGTPDKPDLPSMKOGFNRMTGNTIDFVYVYVYTPQAGTALD 61
 DB 25 APMNRRTPGTEI-----TVEPDDPRYDPLVGNHPRFTG-KPERIHIASSADVHAVA 77
 QY 62 RAMEKSPGTVIVSGGCHYEDFVPECVKAIINVTGLVSGYDDDRGF-VSSGDTNMG 120
 DB 78 DAVR--TGRVVGSRSGCHGFENLVADPAIRLVLDSELRNVYDSTRGAFAIEGALGQ 135
 QY 121 SFKTLFRDGRVLPVGGSCYVGLGHIIVGGDGIILARLHGLPVDMLSGVENVKPVLTED 180
 DB 136 VRTILFKMNGVITITGACPGVAGGHIILGGYGLPSRRFSVVDYLGVEVVDVDADEV 195
 QY 181 SVLKVVHKSDEGDELFWAHTGGGGNGFIITKYYFKD-----LPMSPRGVIA 229
 DB 196 HIVE-ADRNSTGAGHDLMAHHTGGGGNGFIVTRFVRLTPDVVSTDAELPRPATVLL 254
 QY 230 SNLHESWDGFTDALODLLTKYFKLARCWKMTVGFQIFHOAAEFVWYLYTSY----- 284
 DB 255 RSFHWPMHLETSQSFVLLQNF-----GMYEOSHAPBESTOLGLFSTLVCARH 302
 QY 285 -----SNDAREVAODRHVHLEADIEQYKTCCEPTKALGGHAGAPPPVRP 330
 DB 303 QAGYVTLNVLHLDGTDPNAERTLAE---HLSAINAOVGT--PAEGLRETLPMV----- 350
 QY 331 RKRHTSKTSYMDETWDYPPYALTEITINGSGPNQSGKYSAYMIKDPDFQIDVWKYLT 390
 DB 351 --RSTQVAG-----AIAE--GGEPMQRTKVAAYLRTGLSEAGLQATVYRRLT 394
 QY 391 EVPDGLTSAMDALLQVDMFGEIHKVWDATAVAQREYITKQYQTYWQEDKDAVNL 450
 DB 395 ----VYGYDNPAAALLLGGYGMANAVASATALAQDSVLKALFVTWMSPEADERHL 449
 QY 451 KWLIDFYEEMYPEYGGVDPNPQVESGKGFEGCYFNPYDVLDNN--WKNKGKGALELYF 508
 DB 450 TWIRGFYREMYAETGGVVPFGTRV-----DGSYINYPDTLADPLMNTSGVAMHDLTY 502
 QY 509 LGNLNRLIKAKWLMDPNEIFTNKOSIPTKPLKEPKQ 544
 DB 503 KONYPRLOAKAKAMPDQNIPOHGLSTIKPPARLSPGQ 538
 RESULT 6
 AAB70952
 ID AAB70952 standard; Protein; 539 AA.
 AC AAB70952;
 XX
 DT 28-AUG-2001 (first entry)
 XX
 DE S. spinosa protein fragment encoded by ORFs, SEQ ID 16.
 XX
 KW Forosamine; trimethylthamose; polyketide synthase; biosynthesis;
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KW macrolide; insecticidal; C-C cyclising enzyme.
 XX
 OS Saccharopolyspora spinosa.
 XX
 PN DE19957268-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 29-NOV-1999; 99DE-1057268.
 XX
 PR 27-AUG-1999; 99DE-1040596.
 XX
 PA (FARB) BAYER AG.
 PI Eberz G, Moehle V, Froede R, Velten R, Salas JA;
 XX MPI; 2001-267102/28.
 DR N-PSDB; AAF86322.
 XX
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

PT recombinant production of insecticidal spinosyns and their derivatives
 XX
 PS Claim 33; Page 119-121; 35app; German.
 CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polyketide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (II); (ii) to generate a library of polyketide synthases; (iii) for
 CC adding forosamine or trimethylthamose to a spinosyn or polyketide
 CC aglycone; and (iv) for recombinant production of the corresponding
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence represents an S. spinosa C-C cyclising enzyme.
 XX
 SQ Sequence 539 AA;
 Query Match 21.2%; Score 635.5; DB 22; Length 539;
 Best Local Similarity 30.2%; Pred. No. 2,5e-52;
 Matches 174; Conservative 91; Mismatches 216; Indels 95; Gaps 16;
 QY 2 AILPQKPGYIVDVNAGTPDKPDLPSMKOGFNRMTGNTIDFVYVYVYTPQAGTALD 61
 DB 25 APMNRRTPGTEI-----TVEPDDPRYDPLVGNHPRFTG-KPERIHIASSADVHAVA 77
 QY 62 RAMEKSPGTVIVSGGCHYEDFVPECVKAIINVTGLVSGYDDDRGF-VSSGDTNMG 120
 DB 78 DAVR--TGRVVGSRSGCHGFENLVADPAIRLVLDSELRNVYDSTRGAFAIEGALGQ 135
 QY 121 SFKTLFRDGRVLPVGGSCYVGLGHIIVGGDGIILARLHGLPVDMLSGVENVKPVLTED 180
 DB 136 VRTILFKMNGVITITGACPGVAGGHIILGGYGLPSRRFSVVDYLGVEVVDVDADEV 195
 QY 181 SVLKVVHKSDEGDELFWAHTGGGGNGFIITKYYFKD-----LPMSPRGVIA 229
 DB 196 HIVE-ADRNSTGAGHDLMAHHTGGGGNGFIVTRFVRLTPDVVSTDAELPRPATVLL 254
 QY 230 SNLHESWDGFTDALODLLTKYFKLARCWKMTVGFQIFHOAAEFVWYLYTSY----- 284
 DB 255 RSFHWPMHLETSQSFVLLQNF-----GMYEOSHAPBESTOLGLFSTLVCARH 302
 QY 285 -----SNDAREVAODRHVHLEADIEQYKTCCEPTKALGGHAGAPPPVRP 330
 DB 303 QAGYVTLNVLHLDGTDPNAERTLAE---HLSAINAOVGT--PAEGLRETLPMV----- 350
 QY 331 RKRHTSKTSYMDETWDYPPYALTEITINGSGPNQSGKYSAYMIKDPDFQIDVWKYLT 390
 DB 351 --RSTQVAG-----AIAE--GGEPMQRTKVAAYLRTGLSEAGLQATVYRRLT 394
 QY 391 EVPDGLTSAMDALLQVDMFGEIHKVWDATAVAQREYITKQYQTYWQEDKDAVNL 450
 DB 395 ----VYGYDNPAAALLLGGYGMANAVASATALAQDSVLKALFVTWMSPEADERHL 449
 QY 451 KWLIDFYEEMYPEYGGVDPNPQVESGKGFEGCYFNPYDVLDNN--WKNKGKGALELYF 508
 DB 450 TWIRGFYREMYAETGGVVPFGTRV-----DGSYINYPDTLADPLMNTSGVAMHDLTY 502
 QY 509 LGNLNRLIKAKWLMDPNEIFTNKOSIPTKPLKEPKQ 544
 DB 503 KONYPRLOAKAKAMPDQNIPOHGLSTIKPPARLSPGQ 538
 RESULT 7
 AAB91135
 ID AAB91135 standard; Protein; 530 AA.
 XX

AC	ABB91135,
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Herbicidally active polypeptide SEQ ID NO 346.
XX	
KX	Herbicidal; plant; agriculture; herbicide.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200210210-A2.
XX	
PD	07-FEB-2002.
XX	
PF	28-AUG-2001; 2001WO-EP09892.
XX	
PR	28-AUG-2001; 2001WO-EP09892.
XX	
PA	(PARB) BAYER AG.
XX	
PI	Tietjen K, Weidler M;
XX	
DR	WPI; 2002-26910/31.
XX	
PT	Identifying plant target proteins for herbicidally active compounds,
XX	comprising aligning and comparing nucleic acid or amino acid sequences
PT	from plant with nucleic acid or amino acid sequences from non-plant
XX	organisms -
PS	Claim 5; SEQ ID NO 346; 261pd + Sequence Listing; English.
XX	
CC	The invention relates to identifying target proteins
CC	(ABB90790-ABB94016) for herbicidally active compounds, comprising
CC	aligning and comparing nucleic acid or amino acid sequences from plant
CC	with nucleic acid or amino acid sequences from non-plant organisms using
CC	suitable search parameters, where plant sequences having an E-value
CC	greater by a factor of 3 than the E-value of most similar non-plant
CC	sequences are selected. The polypeptides or nucleic acids encoding them
CC	are useful for identifying modulators. The identified modulators are
CC	useful as herbicides.
XX	
SEQ	Sequence 530 AA:
Query Match	12.3%; Score 368.5; DB 23; Length 530;
Best Local Similarity	27.0%; Pred. No. 1.8e-26;
Matches 134; Conservative	76; Mismatches 180; Indels 107; Gaps 23;
QY	72 VRIVGCGHCEPFDVECYKALII-----NTGLVSGSYDDRGVFSSGDTMNGSFKT 124
DB	103 LIRIRGGHDYEELSTWSSVPFILDMYLIRSTIVDVS----KKAWIQAGAT-LGELYT 156
QY	125 LFRDGRVL--PGGSCTSVGLGHIITVGGDGILARLHGSLPVMLSGEVEVVVPVLTEDSV 182
DB	157 NVNDVTSQTLPFGAVCATYGAAGHLSGGYSGLMKRYGTIVDHVIDAQII-----DVN 209
QY	183 LKYVKHSDSENGDELFWAHTGGGGGNFGIIT--KYYFKDLSPSPRGVIASNLHFSWDGFT 240
DB	210 GLTLNRATWGED--LFWAIRGGGGSGFGVILSMKINLVDPKIVT-VFKVNKTLEGG-- 264
QY	241 RALODLLTKYIKLARCDPKMTVGRF--QIFQAABE-----FWMLTYISYSD 287
DB	265 ----TDVLTK-----WQLVASKFPESLFLRANFPQVANGTKRGERTITTVFYAFQIGR 312
QY	288 AEREVAODHYHLLEADIEOIYKTCSEPTKLGGHAGAPRPVRPRKRHTSKTSYMHDETWD 347
DB	313 TDALAIAIMQNMPBELGK--HEDCQMSTLNSTLFPAVDIPA-----GIPTIL-----LD 360
QY	348 YEFVALTETINGSGPNOGRKYSAIWKDFPDFQIDVIWKYLEVPDGLTSAEMKDALLQ 407
DB	361 RP-----SSPDFFKSXDYVKCPDKPREGLEKLMKTMLEKNNNI-----VMWQ 403
QY	408 VMFGEGLHKVWMDATAVAQR-EYIKLOIOTYWQEDDAVNLKIRIPFEAMTEPYGG 466

Db	404	FNPIYGVWDRILPATATAPPHRKGULFKIQYFTTFWNNANAMSSLSIQMKELY-EVAEPY--	460
Qy	467	VPDENTYEGSKGVFECCYFNYPVVDLNMKNKG-----YCALBYFLGNLRPLKA	518
Db	461	-----VSSNP---REAFNRYRIDIVGNSNPGETNVDEAKTYGS--KYLGNLKRIMDV	508
Qy	519	KMLMDPNEIFTNKOSIP	535
Db	509	KAKYDPDNFFKNESQIP	525
RESULT 8			
ABB93022	ABB93022 standard; Protein: 539 AA.		
XX	ABB93022;		
AC			
XX			
DT	31-MAY-2002	(first entry)	
XX			
DE	Herbicideally active polypeptide SEQ ID NO 2233.		
XX	Herbicideal; plant; agriculture; herbicide.		
XX	Arabidopsis thaliana.		
XX	WO200210210-A2.		
PN			
XX	07-FEB-2002.		
PD			
XX	28-AUG-2001; 2001WO-EP09892.		
PF			
XX	28-AUG-2001; 2001WO-EP09892.		
PR			
XX	(FARB) BAYER AG.		
PA			
XX	Tietjen K, Weidler M;		
PI			
XX	WPI; 2002-269010/31.		
DR			
XX			
PT	Identifying plant target proteins for herbicideally active compounds,		
PR	comprising aligning and comparing nucleic acid or amino acid sequences		
PT	from plant with nucleic acid or amino acid sequences from non-plant		
PT	organisms -		
XX			
PS	Claim 5; SEQ ID NO 2233; 261pp + Sequence Listing; English.		
XX			
CC	The invention relates to identifying target proteins		
CC	(ABB90790-ABB94016) for herbicideally active compounds, comprising		
CC	aligning and comparing nucleic acid or amino acid sequences from plant		
CC	with nucleic acid or amino acid sequences from non-plant organisms using		
CC	suitable search parameters, where plant sequences having an E-value		
CC	greater by a factor of 3 than the E-value of most similar non-plant		
CC	sequences are selected. The polypeptides or nucleic acids encoding them		
CC	are useful for identifying modulators. The identified modulators are		
CC	useful as herbicides.		
XX			
XX			
XX	Sequence	539 AA;	
Qy	Query Match	11.6%; Score 347.5; DB 23; Length 539;	
Db	Best Local Similarity	26.2%; Pred. No. 2e-24;	
Db	Matches	141; Conservative 76; Mismatches 181; Indels 141; Gaps 28	
Qy	52	TPQGACTALDR-----AMEKCSFG--TVRIVSGHCYE--DFVFDECVKAIINVTGLV	100
Db	83	TPKPAIIVTPRSDIHSAVAVTCSSKSLNFLKIRISGSHDYEGLSYISDKPF-FILDMSNLR	141
Qy	101	ESGVD-DRGRFVSSGDT-----NMGSFRTLPRDHRVLPGGSCYVVGGHVGGSD	152
Db	142	DVSVDIADQOSAMISAGATLGEVYRIWEKSKY---HG--FPAGVCPVVGSHISGGGY	195
Qy	153	GILARLHGLPVDWISGVVWVVKPVLTEDSVLKYYHKDESGNDGELFWAHTGGGNGFGLI	212
Db	196	GMLLRKFGELSDNLTDAKIV-----DVNGQILDRKSMGED--LFWAISGGGASFGV	246

XX Claim 5; SEQ ID NO 345; 261pp + Sequence Listing; English.

XX
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

XX
XX
SQ Sequence 535 AA;

Query Match 11.0%; Score 328.5; DB 23; Length 535;

Best Local Similarity 27.5%; Pred. No. 1.4e-22; Indels 117; Gaps 25;

Matches 138; Conservative 66; Mismatches 180;

QY 72 VRIVSGHCEYDFVDECVKAI-----NVTGLVESGYDDRGYFVSSGDTNWSFKT 124

DB 103 IRISSGHNDNGLSVSSVFPVILDMHKLRTVDVS-----KKAVVQAGAT-LGELVY 156

QY 125 LFRDHGRLV--PGSSCYSVGLGHIYGGGGITLRLHGLPYDMLSGVEVVVKKVLTEDSV 182

DB 157 KIDBASOTLAFPAICATVAGAGHISGGYGNLMRKFGTTVDHVIDALV-----DVN 209

QY 183 LKVVHKSQNDGELFWAHTGGGNGFCITTKYFKDLPMSRR--GYIASNLHSMWGF 240

DB 210 GKLNRSITMGED--LFWAIRGGGASFGVILSMKI--NLVEVPKLFYFQVVKTLGGG-- 264

QY 241 RDALQDLTKYFKLARCQMKNTVGK--QIFHOAAEFVWLYTSYNDAREVA----- 293

DB 265 ----TDVYK-----WQLVANKFPDNLFLRAMPQV-----NGKRGERTAIAYFWA 307

QY 294 --QDRHYHLEADIQIY-----KTEPTKALGGHAGMAPFVRPRKRTSKTSYHDE 344

DB 308 QFLRTDELIMINQSPFELGLRREDCQMSMLTTLFWAMLPA-----GTPKTVLLGRP 362

QY 345 TMDVPRYALTTETINGSQNGKYSAMIKDPDPFOIDVWKYLTVEPDLGSAEMKA 404

DB 363 T-DVDFP-----KSKSDVKKPIFPEGLEKIWK-----TWLKNNTI 397

QY 405 L-LQVDFEGEIHKVWDATAVAORE-YIILQYQTYQOEEDKDAVNLKIRDFYEEYE 462

DB 398 VMLHFNPGYGMMDLIPSNATLFPHRKGNLFVQYVYTTMLDPNATESLSTIKELY- EYAE 456

QY 463 PYGGVDPDPPTQVESGKGVFEGCYFNYPDVDLNWKNGK-----YGALELYFLGILNR 514

DB 457 PY-----VSSNP--REAFPNYRIDIGSNPSETDVDEAKIYG--YKYFLGNLKR 502

QY 515 LIAKMWLMDPEITNKOSIP 535

DB 503 LMDVKAQSDPENFPKNEOSIP 523

RESULT 11

ABB90805

ID ABB90805 standard; Protein, 541 AA.

XX ABB90805;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 16.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidler M;

XX WPI; 2002-269010/31.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Claim 5; SEQ ID NO 16; 261pp + Sequence Listing; English.

XX
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

XX
XX
SQ Sequence 541 AA;

Query Match 10.8%; Score 324.5; DB 23; Length 541;

Best Local Similarity 25.4%; Pred. No. 3.4e-22;

Matches 132; Conservative 73; Mismatches 169; Indels 145; Gaps 26;

QY 76 SGCHCEYD-----FVDECVKAIINVTGLVESGYDDRGYFVSSGDT-----NW 119

DB 110 SGGHDDGVSYISNRPFVLD--MSYLRNIT--VDVS--DDGSAWGAAGATLGEVYNTIM 164

QY 120 GSEFTLFRDHG--RVLPSSCYSVGLGHIYGGGGITLRLHGLPYDMLSGVEVVVKKVLT 178

DB 165 QSKRT---HGTGFPAGVCPTVAGAGHISGGYGMIMKRYGLSVYVYDAKIV----- 214

QY 179 EDGLVKYHKSQNDGELFWAHTGGGNGFCITTKYFKDLPMSRQVI----- 228

DB 215 -DVNGRIIDRKSMGEB--LFWAIGGGGASFGVILFXIKLVFVPPRVTVFRVEKTLVEN 271

QY 229 ASNLHFSW-----DETRDALDILTKYFKLARCQMKNTVGK--QIFHOAAEFVWY 279

DB 272 ALDMVHKQFVADKTSDDLFRMLQPVV-----RNTT-----QTVRASVVA 313

QY 280 LYSYNDAREEVAODRHYHLEADIQIYKTEPTKALGGHAGWA-----PVPVRPR--K 332

DB 314 LFLGKOSDLMSLTKE---FPELGLRP--ENCTEMTWISVMMANNANDATVTKPILLD 368

QY 333 RHTSKTSYMHDETMVPRYALTTETINGSQNGKYSAMIKDPDPFOIDVIMKYLTVEV 392

DB 369 RNPDSASF-----FNPYGGKMSVATTAAPPFRKRLFKVQHSMMNKKPCTDVSSSF 453

QY 404 GKLGIV-----FNPYGGKMSVATTAAPPFRKRLFKVQHSMMNKKPCTDVSSSF 453

DB 450 LKMWIRDFYEEYEPYGGVDPDPPTQVESGKGVFEGCYFNYPDVDL-----NNMKNGK-Y 501

QY 454 MEKTRSFYSYM-APFYTKNPPRHT-----YLNTRDDLDIGNSGPNRYRAEYV 500

DB 502 GALELYFLGILNLIRAKWLMDEPNEIFTNKOSIPTKPLK 540

QY 501 G--RKYFGENFRLVAVKTAVDPENFRDQSIPTLPK 537

RESULT 12

ABB93021

XX

XX

XX

XX

XX

XX

XX

XX

XX

[illegible]

OY		397	TSAMKQALLQVDMEGGEIHKVVWDATAVAHQREYIIKLQVTWQGE-----DKDAVNLKM	452
Dd		411	V-----FNDYGGKMAEVAVANAPFPFRNKLFIQYSVNKKENSAETKEGYLQAOK	460
OY		453	IRDFEEWEYEPYGVPDPTQVBSGKGVEGCGYPNPVDVL-----NNMKNGK-YGAL	504
Dd		461	V-----LYSMTCGFVSKNRP-----SSRYNYMDVDVIGVNDHGANSYKEGEVVG--	503
OY		505	ELYFLGNILRLIKAKWLMDPNEIFTNKOSIPT	536
Dd		504	RKYFGENFDRLVKIKTAVDPGNFFRNFQSIPIT	535
 RESULT 13 AAy88391 standard; Protein; 545 AA.				
XX	ID	AAy88391		
AC	AAy88391;			
XX	27-JUL-2000	(first entry)		
DE	Tetrahydrocannabinolic acid synthase amino acid sequence.			
KM	Tetrahydrocannabinolic acid synthase; THCA; anaesthesia; pain-killer;			
XX	intracellular pressure lowering; anti-inflammatory treatment.			
OS	Cannabis sativa.			
PN	JP2000078979-A.			
XX	21-MAR-2000.			
PD	04-SEP-1998;	98JP-0251667.		
PF	04-SEP-1998;	98JP-0251667.		
PR	04-SEP-1998;	98JP-0251667.		
XX	(TAIS) TAISHO PHARM CO LTD.			
PA	WIPI; 2000-285930/25.			
DR	N-PSDB; AAA13340, AAA13341.			
XX	New tetrahydrocannabinolic acid synthase gene -			
PT	Claim 2; Page 8-10; 17pp; Japanese.			
PS	This sequence represents the Cannabis sativa tetrahydrocannabinolic acid			
CC	(THCA) synthase amino acid sequence. The THCA gene sequence can be used			
CC	to produce an expression vector containing the gene. Cells can be			
CC	transformed using the vector, so that they produce a protein with THCA			
CC	synthase activity. THCA can be used for anaesthesia, pain-killing,			
CC	intracellular pressure lowering and anti-inflammatory treatment.			
XX	Sequence	545 AA;		
SQ	Query Match	10.2%; Score 306.5; DB 21; Length 545;		
	Best Local Similarity	24.6%; Pred. No. 1.9e-20;		
	Matches 139; Conservative	65; Mismatches 207; Indels 153; Gaps 25;		
OY	36 NRRRIGINIDPVVYVTPPGACTLDADMEKSCSFGT---VRIVSGHCYEDFYDECYKA	92		
Dd	70 NLRFISDTTPKPLVIVPSN--NSHIDAATLICSKKVALQIRTRSGGDHAEQMSYISGVPF	127		
OY	93 II-----NVTTGLVESGYDDRGYFVSSGDTWMSGFKTLFRDHGRVLPQGASC	138		
Dd	128 VVVDLRMNHSTKIDVHSQTAMVEKGALLEGYY-----WINRK----KENLSFGGYC	176		
OY	139 YSVGLGSHIVGGDGILLARLHGLEFVDMLSGEVVVVKPELVLTEDSVLKXVHKDSGNDGELF	198		
Dd	177 PTVGVGGHFSGGGYGALMRNYGLAAD-----NIIDAHVNVWG--KVLDERSMGED--LF	227		
OY	199 WAHTGGGGGNGPGLITTKTYFFDLPMSPRGVIAS-----NLHFSWNGFTRDALODL	247		
Dd	228 WAIRGGGGGENFGIILAANKIKLVAAVPSKSTIFSVMKNMEIHGLVLEFNKKMQNIAYKKYDKDL	287		

XX
PS Claim 5; SEQ ID NO 2231; 261pp + Sequence Listing; English
YY

CC The invention relates to identifying target proteins
CC (AB99790-AB94016) for herbicidally active compounds, comprising
CC attaining and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

SQ Sequence 532 AA;

Query Match	Score	DB	Length
10.0%	300	23	532
33.3%	7	20	

Best local similarity 23.54; Pred. NO. 7.9e-20;
Matches 129; Conservative 78; Mismatches 196; Indels 146; Gaps 24;

```

QY 36 NRMNIGTIDFVVVVYVYVQGACTALDRAHEKSPG---TVRIIVSGHCE-----DF 84
Db 73 NQRSAPRVPKPVLLITVPQSDV---QSAVKCARRRPGHIRTRSGHDYGLSYTHKPE 130
QY 85 VFDECVAKIINVGTGLVESGYD--DDBGYFVSSGDITWNGSFKTLFRDHGR-----VLRGSC 138
Db 131 V-----ILDLRLNRSITVDVDRKNSVMWQGTAT---IGELYELIGKKNRTLAPAGVC 179
QY 139 YSVGLGCHIVGGGDSGLARLHGLPVDWLSGVEVVVYKPVLTEDSVLKYHKDSENGDELF 198
Db 180 PTVAGGGHFGSGGYGTLRKGLADHDVLDARVV-----DARGILRRREMGD--FF 230
QY 199 WAHNGGGGNGGILITKYXFKDLPMSPRGVIAENLHPS-----MGPTRDLQD---246
Db 231 WAINGGGSSSCVLSMKIGLINVSTYTVFVNTYFSQSALKIHRWO--FAADKSDDL 289
QY 247 ---LITKYFLIARCDMKNVTKGFOIFQAAEEFMYILTSYSNDAEREVAODRHYLEA 302
Db 290 FIRWALQRYKMMVNASPFG-----LYL--GSYKNLLKWNKKEPELGLEB 332
QY 303 D-----IEOIYKTCPTTKALGHNAGMAPFVPRPKRHTSKTSYMHDETMQPYFALTE 355
Db 333 DDCTEMSNIEISVIFPAE---LG-----BEPVIVLTK 360
QY 356 TINSGPQRKYSAYAMIKDPDPDQIDVIMKYLEPVRPGLTSAEMKDALQVDMGGERI 415
Db 361 RTRAS---LAFKASDFVQEPMPKTAISKMBRLQD-----PEKHQALLFTPPGGKK 410
QY 416 HKVVMDATAVAQRE--YIITQYQTVAQSEEDKQAVNLKWRDYEEMVEYEPGVDPDENTQV 474
Db 411 SEINDYELPPFHRKKNITYEIQVLANWNRGDVKEKY--MRWRERYDDDMSEVVAASP-----463
QY 475 ESGKGVBEGCYFNYPVDVL-----NNMKNGKYGALELYFLGNINRLIKAKWIDPNE 526
Db 464 -----RGAYINLRDLDLGMVYGVKRSYEEGKSGMWVK--YFKNNPERLVARVXTSVDPSD 515
QY 527 IFTNKOSIP 535
Db 516 FFCDEQOSIP 524

```

```
Search completed: June 10, 2003, 09:58:44
Job time : 45 secs
```

